# LECTIN-BASED PROTEOMIC STUDIES OF URINARY PROTEINS FROM PATIENTS WITH ENDOMETRIAL, OVARIAN AND CERVICAL CANCER

**ALAN MU KANG WAI**

**THESIS SUBMITTED IN FULFILMENT**

**OF THE REQUIREMENTS**

**FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**

**INSTITUTE OF BIOLOGICAL SCIENCES**

**FACULTY OF SCIENCE**

**UNIVERSITY OF MALAYA**

**KUALA LUMPUR**

**2014**

**ABSTRACT**

Endometrial (ECa), ovarian (OCa) and cervical (CCa) cancers are among ten of the most common cancers affecting women worldwide. Cancers are known to cause some proteins to be differentially glycosylated or aberrantly excreted in the urine, which can be used as biomarkers. Since ECa, OCa and CCa are difficult to diagnose at the early stage, the aim of the present study was to identify a panel of new complementary biomarkers for early detection of the cancers. Identification of early biomarkers which are specific and efficient can increase the survival rate of the patients.

To screen for potential biomarkers for ECa, separation of urinary proteins of control subjects and patients with ECa by 2-DE and densitometry analysis were performed. The analysis of 2-DE urinary profiles of controls and patients with ECa demonstrated significant differential expression of zinc-alpha 2 glycoprotein (ZAG), acid alpha-1 glycoprotein (AAG) and CD59. When similar 2-DE analysis was performed using CGB lectin-Western blotting approach to detect *O*-glycosylated urinary proteins, the level of nebulin (NEB) was detected to be reduced in the ECa patients compared to the controls. Using similar analysis with CGB lectin, the altered levels of kininogen (KNG), clusterin (CLU) and leucine rich alpha-2 glycoprotein (LRG) were also demonstrated in the patients with OCa. However, the expression levels of all *O*-glycosylated proteins appeared to be comparable in patients with CCa.

A significant increase in the levels of urinary ZAG and AAG was detected in ECa patients when the 2-DE profiles were transferred onto membrane and probed with CMB lectin to exclusively capture *N*-glycosylated urinary proteins. Similar comparative analysis of the *N*-glycosylated protein profiles performed on OCa patients demonstrated overexpression of AAG, ZAG, LRG, CLU and HPβ chain but decreased

in the levels of KNG. Comparable expression of *N*-glycosylated urinary proteins was detected between the controls and CCa patients.

Analysis of bound fractions obtained from immoblilized CGB lectin affinity chromatography followed by the labeled-free quantification showed differential levels of LRG and NEB in patients with OCa and ECa, respectively. This is comparable with the earlier data obtained from CGB lectin probed Western blot. Similarly, spectral counting of the *N*-glycosylated proteins isolated by immobilized CMB lectin affinity chromatography demonstrated altered levels of AAG and ZAG for patients with ECa and OCa and differential expression of KNG and HPβ chain for OCa patients, which are also compatible with the findings using CMB lectin blot. In addition, the LC-MS/MS analysis of the CMB lectin bound fractions detected CD59 peptides only in the controls but not in patients with ECa, OCa and CCa.

When SELDI-TOF and Biomarker Wizard analyses were carried out on the CMB lectin captured urinary glycopeptides from control subjects and patients with ECa, OCa and CCa, peaks *m/z* 1201 and 1449 were detected as potential group discriminators. Similarly, the analysis performed on the CGB lectin captured urinary glycopeptides from control subjects and the three groups of cancer patients predicted peaks *m/z* 2138 and 3656 as potential group discriminators.

Taken together, the findings of this study suggest urinary proteins such as ZAG, AAG, NEB, KNG, LRG, CLU, HPβ chain and CD59 may serve as potential complementary biomarkers for the early detection of ECa, OCa and CCa although this requires further extensive validation on clinically representative populations.

**ABSTRAK**

Kanser endometrium (ECa), ovari (OCa) dan serviks (CCa) adalah di antara sepuluh kanser yang paling lazim melibatkan wanita di seluruh dunia. Kanser sememangnya diketahui menyebabkan sesetengah protein mengalami perubahan pengglikosilan atau dikumuhkan melalui air kencing, yang boleh diguna sebagai penanda biologi. Disebabkan ECa, OCa dan CCa adalah sukar untuk didiagnosis pada peringkat awal, tujuan kajian ini adalah untuk mengenalpasti satu panel penanda biologi pelengkap baru untuk mengesan kanser pada peringkat awal. Pengenalpastian penanda biologi baru yang spesifik dan berkesan boleh meningkatkan kadar kemandirian pesakit.

Untuk menyaring penanda biologi yang berpotensi bagi ECa, pengasingan protein air kencing daripada subjek kawalan dan pesakit ECa dilaksanakan dengan menggunakan kaedah 2-DE dan analisis densitometry. Analisis menunjukkan perbezaan ketara dalam pengekspresian zink-alfa 2 glikoprotein (ZAG), asid alfa-1 glikoprotein (AAG) dan CD59. Apabila kaedah 2-DE yang sama dilakukan dengan menggunakan kaedah pemblotan Western-lektin CGB untuk mengesan protein air kencing yang *O*-terglikosilat, paras nebulin (NEB) dikesan berkurangan dalam pesakit ECa berbanding subjek kawalan. Dengan menggunakan kaedah yang sama dengan lektin CGB, penjejasan pengekspresian kininogen (KNG), klusterin (CLU) dan glikoprotein kaya leusina alfa-2 (LRG) turut ditunjukkan berubah dalam kajian yang sama. Walau bagaimanapun, tahap pengekspresian kesemua protein *O*-terglikosilat kelihatan setanding dalam pesakit CCa.

Peningkatkan aras ZAG dan AAG air kencing dikesan berlaku dalam pesakit ECa apabila profil 2-DE dipindahkan ke atas membran dan disurih dengan lektin CMB untuk mengesan protein *N*-terglikosilat secara eksklusif. Analisis sama yang dilakukan

pada profil protein *N*-terglikosilat pesakit OCa menunjukkan peningkatan pengekspresian protein AAG, ZAG, LRG, CLU dan HPβ tetapi penurunan aras KNG. Walau bagaimana pun, aras pengekspresian kesemua protein air kencing *N*-terglikosilat adalah setanding bagi pesakit CCa.

Analisis pecahan berikat yang diperoleh daripada kromatografi keafinan lektin CGB diikuti dengan kuantifikasi bebas label menunjukkan pembezaan aras LRG dan NEB masing-masing dalam pesakit OCa dan ECa. Ini adalah serasi dengan penemuan yang dihasil menggunakan "Western blot" yang disurih lektin CGB. Begitu juga, pengiraan spektrum protein *N*-terglikosilat yang diasing menggunakan kromatografi keafinan lektin CMB menunjukkan perubahan ketara aras AAG dan ZAG dalam pesakit ECa dan OCa dan perbezaan pengekspresian KNG dan HPβ dalam pesakit OCa, yang juga serasi dengan penemuan menggunakan blot lektin CMB. Selain itu, analisis LC-MS/MS terhadap pecahan berikat lektin CMB mengesan kewujudan CD59 hanya di dalam subjek kawalan dan tidak terdapat dalam pecahan yang dikumpul daripada pesakit ECa, OCa dan CCa.

Apabila analisis SELDI-TOF dan Biomarker Wizard dijalankan terhadap glikopeptida air kencing pesakit ECa, OCa dan CCa yang terikat kepada lektin CMB, puncak *m/z* 1201 dan 1449 dikesan sebagai kumpulan pembeza yang berpotensi. Begitu juga, analisis yang dilakukan ke atas glikopeptida yang diikat oleh lektin CGB daripada subjek kawalan dan tiga kumpulan pesakit kanser turut meramalkan *m/z* 2138 dan 3656 sebagai kumpulan pembeza yang berpotensi.

Apabila dikumpulkan, hasil kajian ini mencadangkan protein air kencing seperti ZAG, AAG, NEB, KNG, LRG, CLU, rantaian HPβ dan CD59 berpotensi untuk

digunakan sebagai penanda biologi awal untuk pengesanan ECa, OCa dan CCa walaupun ini memerlukan pengesahan selanjutnya dalam populasi klinikal.

**Acknowledgement**

First of all, I would like to express heartfelt thanks and deep appreciation to my supervisors, Dr Adawiyah Suriza Shuib and Prof Dr Onn Hashim, who have guided and advised me throughout the study. Without their tireless help and valuable suggestions, this study will not be accomplished.

Special thanks to my labmates Suhana, Shahirah, Joe Lau, Amy Tan, Lum Ai Theng for their kindness in helping me whenever I needed. I truly appreciated to their sacrifice and disinterested supports. For those whose names are not mentioned, all your guidance and support will be remembered.

Last but not least, I would like to express my sincere gratitude to my parents, family members and friends who have provided constant moral support and assistance.

## CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

**LIST OF ABBREVIATIONS**

| | |
|---|---|
| 2-DE | 2-dimension electrophoresis |
| AAG | Alpha-1 acid glycoprotein |
| AMBP | Protein AMBP |
| AP | Alkaline phosphatase |
| APS | Ammonium persulfate |
| APOD | Apolipoprotein D |
| ATR | Antithrombin III |
| BSA | Bovine serum albumin |
| CCa | Cervical cancer |
| CGB | Champedak galactose binding |
| CHCA | α-cyano-4-hydroxycinnamic acid |
| CLU | Clusterin |
| CMB | Champedak mannose binding |
| CNBr | Cyanobromide |
| Con A | Concanavalin A |
| DVS | Divinylsulfone |
| DTT | Dithioreitol |
| ECa | Endometrial cancer |
| EDM | Expression Different Mapping |
| EPO | Erythropoietin |
| FIGO | Federation of International Gynaecological and Obstetricians |
| FT-ICR | Fourier transform-ion cyclotron resonance |
| Gal | Galactose |
| GalNAc | N-acetylgalactosamine |
| Glc | Glucose |
| GlcNAc | N-acetylglucosamine |
| HPLC | High performance liquid chromatography |
| HPV | Human papillomavirus |
| HRP | Horseradish peroxidase |
| HP | Haptoglobin |
| HPβ | Haptoglobin β chain |
| HSPG2 | Heparan sulfate proteoglycan 2 |
| IARC | International Agency for Research on Cancer |
| IGHA1 | Ig alpha-1 chain C region |
| IGHG3 | Ig gamma 3 chain C region |
| IL-12 | Interleukin-12 alpha chain |
| IPI | International Protein Index |
| ITIH4 | Inter alpha trypsin inhibitor heavy chain H4 |
| KAC | Ig kappa chain C region |
| KNG | Kininogen-1 |
| LC | Liquid chromatography |
| L-fuc | L-fucose |
| LRG | Leucine rich alpha-2 glycoprotein |
| MALDI | Matrix-assisted laser desorption/ionization |
| Man | Mannose |

| | |
|---|---|
| MS/MS | Tandem mass spectrometry |
| *m/z* | Mass over charge |
| NEB | Nebulin |
| NSAF | Normalized spectral abundance factors |
| OCa | Ovarian cancer |
| PBS | Phosphate buffered saline |
| PHA | Phytohemagglutinin |
| p*I* | Isoelectric point |
| Q | Quadrupole |
| RBP4 | Retinol binding protein 4 |
| SBA | Soybean agglutinin |
| SDS-PAGE | Sodium dodecyl sulphate-polyacrylamide gel electrophoresis |
| SELDI | Surface enhanced laser desorption ionization |
| S.E.M | Standard error mean |
| S/N | Signal to noise |
| TAGLN | Transgelin |
| TEMED | N,N,N’,N’-tetramethylethylenediamine |
| TBS | Tris buffered saline |
| TBST | Tris buffered saline Tween |
| TOF | Time of flight |
| v/v | Volume over volume |
| WGA | Wheat germ agglutinin |
| w/v | Weight over volume |
| xg | Acceleration due to gravity |
| ZAG | Zinc alpha-2 glycoprotein |

# Chapter 1
# Introduction

# CHAPTER 1: INTRODUCTION

## 1.1 Lectins

### 1.1.1 History of lectin

Lectin was first discovered by Hermann Stillmark in 1888 when he described the agglutination properties of ricin that was extracted from the seeds of castor bean (*Ricinus communis*) of the family of Euphorbaceace (Damme 2008). Later in 1919, the lectin research gained momentum when James B. Sumner was able to purify and crystallize hemagglutinin from Jack bean (*Canavalia ensiformis*), which he named concanavalin A (Con A) Sumner found that Con A can bind and agglutinate red blood cells and precipitate glycogen, mucins and starch (Sumner and Howell, 1936). In 1954, Boyd and Shapleigh introduced the term "lectin" which was derived from the Latin word "legere", meaning "to select" or "to gather" (Hamid *et al.,* 2013). They identified that the lectin extracted from lima bean (*Phaseolus lunatus limensis*) was able to bind and agglutinate human red blood cells of the A-type blood group.

Initially, lectin is defined as a sugar binding protein of non-immune origin which is able to agglutinate cells or precipitate glycoconjugates (Goldstein *et al.,* 1980) However, this definition was expanded to a more concise description for lectins as more lectins with different properties and capabilities were discovered. Therefore, lectin is later denoted as a protein that are non-immunoglobulins capable of specific recognition and reversible binding to carbohydrate moieties of complex carbohydrate without altering the covalent structure of any of the recognized glycosyl ligands (Kocourek and Horejsi, 1983). Currently, a simple definition is used to describe lectin as a carbohydrate-binding protein other than an enzyme or an antibody.

### 1.1.2 Plant lectins: function, localization and classification

Lectins can be found in a variety of life ranging from microorganisms such as viruses and bacteria to plants and animals. They have similar functions regardless of their origin. Additionally, they share a similar binding property toward carbohydrate although they show no homology in primary structure. Through structural elucidation, their binding property was found to be contributed by their tertiary structure, carbohydrate-recognition domain by recognizing the terminal non-reducing carbohydrates residues and the anomeric isomers of glycoproteins (Ghazarian *et al.*, 2011). Generally, they are ubiquitous in plants, especially from the genus of Euphorbiaceae and Leguminoseae. They can be found in roots, tubers, bulbs, leaves, flowers, rhizomes, fruits, stems and seeds. Among these sources, it is more abundant in the seeds, the storage organs of plants.

At present, there are three approaches to classify plant lectins. The first classification has been adopted based on the origin, structurally and evolutionally related to the plant lectin. Under this classification, lectins can be divided into seven distinct families, which include Amaranthin lectins, chitin-binding lectins, Cucurbitaceae phloem lectins, jacalin related lectins, legume lectins, monocot mannose-binding lectins and type-2 ribosome-inactivating proteins (Van Damme *et al.*, 2004).

The second classification method involves the division of lectins into four classes according to their physical structure. The four classes are merolectin, hololectin, chimerolectin and superlectin (Vandenborre *et al.*, 2011). Merolectins are monovalent proteins that consist of a single carbohydrate binding domain. Due to their monovalent property, they are unable to precipitate glycoconjugates or agglutinate cells. Hololectins are lectins that contain at least two homologous or identical carbohydrate-

binding domains. They are able to bind and precipitate glycans because of their multiple binding sites. The third class of lectins is chimerolectins which refers to those proteins that contain a carbohydrate-binding domain arrayed tandemly with an unrelated domain. Their unrelated domain contributes to the catalytic activity that acts independently on the carbohydrate binding domain. Superlectins consist of two carbohydrate-binding domains that recognize structurally unrelated sugars. They can be considered as a special type of chimerolectins composed of two tandemly arrayed structurally and functionally different carbohydrate-binding domains.

The last method of classification is based on lectin's complementary to the carbohydrate groups. Different lectins have their own specificity to the carbohydrate groups. Therefore, plant lectins are divided into five subgroups according to their specificity to the monosaccharide. They are glucose and/or mannose (Glc and/or Man) - specific, galactose and/or *N*-acetylgalactosamine (Gal and/or GalNAc)-specific, *N*-acetylglucosamine (GlcNAc)-specific, L-fucose (L-Fuc)-specific and sialic acid (Goldstein and Poretz, 1986).

Owing to their capabilities of recognizing diverse sugar structures, they mediate a variety of biological processes. One of the most important functions exhibited by plant lectins is to serve as plant storage proteins as they are abundant in seeds and vegetative storage tissues (Van Damme *et al.,* 1995). Besides, it also functions as a protector to defend against phytopathogens (virus, bacteria and fungi) and insects (Peumans and Van Damme 1995). It defends against these invasions through indirect mechanisms, which involved interactions with the cell wall carbohydrate or extracellular glycans (Peumans and Van Damme 1995). Lectin is also involved in

rhizobium-legume interaction, in which the fixation of nitrogen occurs through the symbiosis interaction of the bacteria and the plant roots (Diaz *et al.,* 1989).

Other than these three main functions, lectins have also been shown to have various other physiological functions such as in the cell wall extension, maintaining seed dormancy, mitogenic stimulation of embryonic plant cells, cell recognition between pollen and stigma or style, transporting marcomolecules and packaging or mobilizing the storage materials during the seeds maturation or germination (Vandenborre *et al.,* 2011).

### 1.1.3 Application of plant lectins

The specificity of lectins to recognize particular carbohydrate structures enables them to be used as a tool for various researches. Table 1.1 shows examples of plant lectins and their affinities. Lectins are easily available because they are abundant in plant seeds. In addition, they are less expensive, better characterized with respect to binding specificity and more stable as compared to antibody (Etzler, 1979). Currently, lectins are extensively used to investigate the biological and pathological significance of glycosylation (Dennis *et al.,* 1999; Ghazarian *et al.,* 2011). Because of their capabilities of differentiating the carbohydrate residues available on the blood cells surface, some plant lectins are being used for blood group typing and cell typing. For instance, lectin isolated from *Phaseolus limensis* is used to identify type A blood cell while lectins purified from *Dolichos biflorus* is able to differentiate A1 blood subgroups (Nakajima *et al.,* 1986; Sharon and Lis 2004). Meanwhile, lectin obtained from *Lotus tetragonolobus* could identify type O blood cells (Renkonen, 1948). Lectins from *Vicia graminea* is used to distinguish between the N blood group and the M blood group as it recognizes

only N blood group (Duk *et al.*, 1981). The H blood group antigen is able to be identified by lectin of *Ulex europaeus* (Engelmann, 1993).

**Table 1.1: Examples of plant lectins and their affinity (Dai *et al.*, 2009)**

| Name/ Abbreviation | Origin | Affinity |
|---|---|---|
| LCA | *Lens culinaris* | Fucα1-6GlcNAc and α-Man, α-Glc |
| PSA | *Pisum sativum* | 1,6-Fucosylation of the trimannosyl core and α-Man. |
| AAL | *Aleuria aurantia* | Fucα1-6-GlcNAc, Fucα1,3/1,4-GlcNAc, Fucα1,2-Gal |
| AAA | *Anguilla Anguilla* | Fuc linked to the GlcNAc |
| LTA | *Lotus tetragonolobus* | Fucα1,3/1,4-GlcNAc, Fucα1,2-Gal, $Le^x$ |
| ConA | *Concanavalin ensiformis* | Two nonsubstituted or C2-substituted α-mannopyranosyl residues in one molecule |
| Jacalin | *Artocarpus integrifolia* | *O*-linked oligosaccharide; α-D-Gal,Gal(β-1,3)GalNAc |

Apart from that, plant lectins' agglutination property enables them to be used as diagnostic probes to identify bacteria. For example, the ability of detecting *Neisseria* and *Bacillus* species by wheat germ agglutinin (WGA) lectin and soybean agglutinin (SBA), respectively, has been utilized in the development of a diagnostic kit to differentiate the respective species within the same genus (Schalla *et al*., 1985).

Plant lectins could also be used in studying and characterizing the structure of glycoconjugates. Aberrant glycosylation structures are usually associated with the

disease progression. In order to examine the alteration of disease-related glycosylation, several analytical methods are utilized including lectin affinity chromatography coupled with mass spectrometry. In general, glycoproteins are isolated from the complex biological samples on the basis of their specific carbohydrare group which could be recognized by their respective lectins. For example, Zhao and his coworkers used three types of lectins, including WGA, Elderberry lectin and *Maackia amurensis* lectin that were immobilized to generate affinity chromatography to purify sialylated glycoproteins from sera of normal subjects and patients with pancreatic cancer (Zhao *et al.,* 2006). This was followed by characterization of the glycoproteins and glycans using the mass spectrometry. In addition, lectin affinity fractionation coupled with 2-dimensional electrophoresis (2-DE) was used to identify different isoforms of prostate cancer antigen in the serum samples of patients with prostate cancer (Sumi *et al.,* 2001).

Lectins have also been shown to have the ability to stimulate the division of cells, especially mononuclear cells. For example, lymphocytes cells divide when stimulated with the mitogenic lectin phytohemagglutinin (Cass, Pham et al.) (Reed *et al.,* 1986). Therefore, they are used in the diagnostic tools for lymphoid tissue diseases. An example of this application has been demonstrated when the acute lymphoblastic leukemia was diagnosed by characterizing the immunological phenotype of malignant lymphoid cells and distinguishing the cytological variants of acute lymphoblastic leukemia and lymphomas of lymphatic cells (Gluzman *et al.,* 1991).

Lectins have also played an important role in biomedical and pharmacological fields. Some of the plant lectins such as abrin, cytotoxic A-chains of ricin and purified mistletoe lectin ML-I have been used in therapeutic agents for cancer because of their

antitumor cytotoxin activity where they can inhibit protein synthesis with high efficacy (Lord *et al.,* 1991; Tonevitsky *et al.,* 1991; Paprocka *et al.,* 1992). They have also been used to serve as carriers in targeted therapy for cancer where they are conjugated to drugs and targeted to specific tumors (Bonfils *et al.,* 1992; Mody *et al.,* 1995) Meanwhile, plants lectins such as *Urtica dioica* agglutinin, *Galanthus nivalis* agglutinin, *Astrocarpus heteophyllus* agglutinin, *Myrianthus holstii* lectin and concanavalin A (ConA) lectin were suggested as effective agents to control the course of AIDS as they demonstrated to have anti-HIV activities, *in vitro* (Witvrouw *et al.,* 2005).

### 1.1.4 Champedak

*Artocarpus integer*, commonly known as champedak in Malaysia, is classified under the Moraceae family and *Artocarpus* genus. Champedak is similar to jackfruit (*Artocarpus heterophillus*/ *integrifolia*), but smaller in size, with less pith, more seeds and softer pulp (Abdul Rahman *et al.,* 2002). The seeds of champedak contain two types of lectins which are known as champedak galactose binding lectin (CGB) and champedak mannose binding (CMB) lectin.

#### 1.1.4.1 Champedak Galactose-binding (CGB) Lectin

Champedak galactose-binding (CGB) lectin recognizes the structure Galβ1-3GalNAc of IgA1 and C1 inhibitor (Hashim *et al.,* 1991). It was demonstrated to have haemagglutinating activity against human erythrocytes of blood groups A, B, AB and O (Abdul Rahman *et al.,* 2002). In addition, it was able to compensate the complement and thus decrease the complement-mediated hemolytic activity of sensitized sheep erythrocytes (Hashim *et al.,* 1994). Currently, the lectin has been used as a probe to

study the serum or urinary glycoproteins of several malignancies (Abdul-Rahman *et al.*, 2007; Mohamed *et al.*, 2008; Jayapalan *et al.*, 2012; Mu *et al.*, 2012; Mu *et al.*, 2013).

The structure and properties of the CGB lectin are generally identical to jacalin, a lectin of jackfruit. However, the CGB lectin from different clones of champedak seeds were demonstrated to be structurally and functionally similar but jacalin from diverse cultivars of jackfruits varied in their structures and functional activities (Hashim *et al.*, 1991). The CGB lectin comprises two types of non-convalently linked subunits with Mr of 13000 and 16000 Da. When it was subjected into reverse-phase high performance liquid chromatography (HPLC), the lectin resolved into two major (α and $\beta_1$) and one minor ($\beta_2$) peaks (Abdul Rahman *et al.*, 2002). The α and β subunits contain 133 and 21 amino acid residues, which are non-identical from each other, respectively (Gabrielsen *et al.*, 2009). The first 47 amino acids residues of the *N*-terminal sequence of the α subunit showed close sequence homology with that of jacalin.

Both the $\beta_1$ and $\beta_2$ subunits have the same number of amino acid residues but different in four residues at positions 6, 10, 12 and 19. They share more than 70% sequence homology with the β subunits of other galactose-binding lectins from other Artocarpus species. Both β subunits lack cystein residue in their sequence leading to the failure of interchain or intrachain disulphide bond formation.

#### 1.1.4.2 Champedak Mannose-binding (CMB) Lectin

Champedak mannose binding (CMB) lectin is a homotetramer with a single monomer with a molecular weight of 16800 Da. It has quite a similar structure with artocarpin (the jackfruit mannose binding lectin) but they were not identical due to the

presence of disulphide linkage in the CMB lectin (Lim *et al.*, 1998). The CMB lectin was shown to interact strongly with human IgE and IgM but interact weakly with IgA2 (Gabrielsen *et al.*, 2010). In addition, it also demonstrated a strong interaction with core-mannosyl residues of the *N*-linked oligosaccharides of glycoproteins such as $\alpha_1$-antitrypsin, haptoglobin β chain, orosomucoid, $\alpha_1$β-glycoprotein, ovalbumin, porcine thyroglobulin, human $\alpha_1$-acid glycoprotein and transferrin (Hashim *et al.*, 2001). Man α1-3 Man ligand is the preference isomer for the CMB lectin to bind to as compared to other two types of isomers (Lim *et al.*, 1998). The lectin was also shown to have a selective mitogenic activity towards T lymphocytes (Lim *et al.*, 1998). It activates murine B cells to secret immunoglobulins, but only in the presence of T cells and macrophages (Lim *et al.*, 1998).

## 1.2 Urinary proteomics

### 1.2.1 Definition of proteomics

Proteomic is the study of protein expression in a tissue or biological fluid (Anderson and Anderson 1998). However, this definition is too general to describe proteomics as many different types of approaches are classified under this group other than the study of protein expression. This includes protein-protein interaction, protein post-translational modification and protein structure (Anderson *et al.*, 2000; Graves and Haystead 2002). Therefore, a more accurate and decisive definition has termed proteomic as a large-scale analytical study of protein properties in obtaining a global, integrated view of disease processes, cellular functions and networks at the protein level (Blackstock and Weir 1999).

### 1.2.2 Overview of urine analysis

Compared to other biofluids, urine has several characteristics that makes it a preferred choice for biomarker discovery. Firstly, high amounts of urine are easily and non-invasively obtained from the volunteers. Secondly, the components of the urine are similar to those found in the blood. In healthy individuals, seventy percent of the urine proteome is originated from the kidney and the urinary tract, while the remaining is the proteins filtered by the glomerulus (Decramer *et al.*, 2008). Approximately 150 liters of plasma is filtered by the glomerulus from a total of 900 liters of plasma flowed through the kidneys. The filtration of the glomerulus is mainly a selective process based on dimension, charge and configuration of the protein. Thus, the urinary proteome can be used to detect not only abnormalities within the kidney and the urogenital tract but also systemic diseases associated with small circulating proteins and peptides that can pass through the glomerulus. Thirdly, the urinary proteins are relatively stable without significant protein degradation. Previous studies have shown that the urinary proteome did not change significantly when urine was stored for three days at $4^{o}$C, 6 hours at room temperature and several years at -20 $^{o}$C (Fliser *et al.*, 2007). Fourthly, the urinary proteome contains minimal amount of high abundance proteins such as albumin and IgG compared to serum proteome. Albumin and IgG sometimes mask the low abundance proteins that may have a potential to be a biomarker. Therefore, this makes the analysis of low abundance proteins becomes much easier in urine samples.

Hermogenes has already used the urine's colour as an indicator for a disease in the $5^{th}$ century (Iorio and Avagliano, 1999). In addition, a physician Gerrit Dou described the meticulous observation of urine by the $17^{th}$ century (Fliser *et al.*, 2007). The first urinary proteomic was applied in analyzing the nature of the normal urinary

proteins by Rigas and Heller in 1951 using SDS-PAGE (Fliser *et al.,* 2007). In 1979, Anderson was the first person who studied the normal urine composition using the 2-DE technique (Candiano *et al.,* 2010). The first extensive optimized human urinary proteome map was established by Thongboonkerd and his co-workers in 2002 when they precipitated the urinary proteins from healthy subjects and analysed with 2-DE (Thongboonkerd *et al.,* 2002). Since then, a number of attempts in defining more advance and complete reference maps of normal urinary proteins were undertaken by other researchers using various types of proteomic techniques (Oh *et al.,* 2004; Pieper *et al.,* 2004; Sun *et al.,* 2005; Casado-Vela *et al.,* 2011; Marimuthu *et al.,* 2011).

### 1.2.3 Application of urinary proteome analysis in cancer

Owing to the origin of urine from the renal system, the proteomic analyses have been applied in the search of biomarkers for renal diseases, such as glomerular disease, acute renal allograft rejection, urological cancers and urolithiasis (Cadieux *et al.,* 2004; Schaub *et al.,* 2004a; Thongboonkerd *et al.,* 2004; Goodison *et al.,* 2009; Morrissey *et al.,* 2010). Urological cancers, including bladder cancer and renal cell carcinoma have been extensively examined in the last decade by using this approach. Recently, Chen et al. (2010) identified a panel of biomarkers for bladder cancer using the iTRAQ-based quantitative profile. Some of the proteins, such as apoliprotein A-I, apoliprotein A-II, heparin cofactor 2 precursor and peroxidase-2, were significantly increased in the urine obtained from patients with bladder cancer. Rogers et al. (2003) defined clinically specific biomarkers for renal cell carcinoma using SELDI-TOF analysis, in one training and two blinded sets of urine samples from patients.

Thirteen percent of the urinary proteins are not originated from the urogenital tract. This is not a surprised since proteomic analysis of the urine has revealed biomarkers for several non-renal cancers. Tantipaiboonwong et al. (2005) analyzed the urine obtained from the patients with lung cancer using SDS-PAGE coupled with HPLC and 2-DE analyses. He identified CD59, transthyretin, GM2 activator protein and Ig-free light chains as differentially expressed proteins which may have a potential to be the panel of biomarkers for lung cancer. In a separate study, capillary electrophoresis coupled with mass spectrometry was used to uncover potential urinary biomarkers for prostate cancer (Theodorescu *et al.,* 2008). The results illustrated that these novel biomarkers were secreted into the urine from the prostatic fluid. Meanwhile, gynaecological cancers have also been investigated for urinary protein biomarkers, particularly ovarian cancer. Using 2-DE coupled with mass spectrometry analysis, Abdullah-Soheimi et al. (2010) uncovered a fragment of inter- alpha-trypsin inhibitor heavy chain 4, a fragment of albumin and kininogen that can be potentially used to identify patients with early stage ovarian cancer. Another study examined the feasibility of using urine as a clinical sample to compare and discriminate benign and malignant ovarian cancer proteomic profiles with equalizer bead technology combined with surface enhanced laser desorption ionization (SELDI-TOF) mass spectrometry analysis (Petri *et al.,* 2009).

## 1.3 Proteomic techniques

The strategies of search for biomarkers of disease using proteomics have been categorized as top-down or bottom-up. The top-down strategy comprises the separation of proteins either with chromatography or 2-DE followed by tryptic digestion and the analysis of tryptic peptides using mass spectrometry. On the other hand, the bottom-up

strategies involve the trypsin digestion of proteins into peptides followed by the isolation of peptides with chromatography and analysis using mass spectrometry. The techniques used to separate and analyze the proteins include 2-DE coupled with mass spectrometry, liquid chromatography coupled to mass spectrometry and SELDI-TOF mass spectrometry.

### 1.3.1 Two-dimensional gel electrophoresis

Although non-gel-based approaches have been developed, 2-DE still remains the most popular and effective technique to study and separate proteins from a sample despite being laborious, time-consuming, hard to automate and not applicable for low molecular weight proteins. This technique separates proteins into two dimensions according to their isoelectric point and molecular mass. The first dimension separation involves the fractionation of the proteins by electrofocusing. The proteins migrate to their respective isoelectric point in a pH gradient. This is followed by the separation of proteins in a perpendicular dimension, in which the proteins migrate on the basis of their molecular weights using SDS-PAGE.

After separation in a 2-DE gel, the proteins are resolved into spot patterns, which is often viewed by staining the gel with various staining methods and subsequently converted to compatible image file for densitometry analysis. The volumes of protein spots are determined by quantification software. The differentially expressed proteins are identified by mass spectrometry. The step involves the proteolytic digestion of the proteins by exposing gel slices to trypsin, the extraction of the proteolytic fragments from the gel slices and the identification of the proteins from the database based on the masses of proteolytic fragments as determined by mass spectrometry.

Post-translational modifications have often displayed a broad range of protein patterns such as multiple spots or smears in 2-DE gels. This is because post-translational modifications confer changes of the isoelectric points and molecular weights of the protein. Therefore, 2-DE proves to be able to resolve proteins that underwent high degree of post-translational modifications. For example, 2-DE employed with lectin blotting and lectin affinity chromatography has been used to isolate and analyze the serum *N*-glycosylated proteins obtained from patients with nasopharyngeal cancer (Seriramalu *et al.,* 2010).

### 1.3.2 Liquid chromatography

Liquid chromatography (LC) is a powerful tool which has been used in various applications including the study of the urinary proteome (Agron *et al.,* 2010; Tan *et al.,* 2010). It has gained popularity nowadays as it provides high sensitivity and selectivity towards the analysis. Using this method, the analytes are initially digested with trypsin to form a complexity of the resulting mixture. The resulting tryptic mixtures are fractionated by LC column and later sequenced by compatible mass spectrometer to identify the proteins based on the sequence information of the peptide detected.

### 1.3.3 Mass spectrometry

Mass spectrometry analysis is the last step of the proteome study after the separation step and ionization step. Generally, a mass spectrometer instrument consists of three modules: an ion source, a mass analyser and a detector. Matrix-assisted laser desorption/ionization (MALDI) is one of the most widely used soft ionization technique for the mass spectrometry analysis. It was first introduced by Franz Hillenkamp in 1985 when he found that the amino acid alanine could be ionized easily

with mixing of amino acid tryptophan (Karas *et al.,* 1985; Laiko *et al.,* 2000). It allows the analysis of the biomolecules because it does not tend to destruct them by vaporization and ionization. In MALDI analysis, a matrix compound is used to co-crytallized with the analytes, which plays the role as the proton donor or receptor and energy absorbing material. The ionization process is triggered when a pulsed laser is shot on the matrix to cause it to excite and directly vaporize with the analyte. Finally, the ionized molecules are desorpted into the gas phase and entered the mass analyzer.

Basically, there are four types of mass analyzer typically used for the proteome analysis: quadrupole (Q), ion-trap, time of flight (TOF) and Fourier transform-ion cyclotron resonance (FT-ICR) instruments (Fliser *et al.,* 2007). The TOF analyzer is normally combined with MALDI instrument. In the mass analyzer, the ions are accelerated by an electric field to "fly" down a flight tube and a detector at the end of the tube records the time of flight of the ions. The time of flight of the ion is proportional to the size of the ion. Therefore, the mass of the ion (*m/z*) corresponds to the time of flight of the ion.

In order to obtain highly accurate sequence information of tryptic peptides, tandem mass spectrometry (MS/MS) is performed with the compilation of MALDI and TOF/TOF instrument. The first TOF instrument serves as a mass filter to collect the parent ions, while the second TOF instrument is used to analyze the daughter ions (fragments) that are generated by collision-induced dissociation or electron transfer dissociation (Fliser *et al.,* 2007).

The resulting peptides fragment masses were presented as MS data and subjected to search by engines such as MASCOT, SEQUEST, ProFound or Spectrum Mill. Currently, there are two frequently used databases which are the Swiss-Prot and the International Protein Index (IPI). The search engine compares the MS data against the database consisting of the amino acid sequence masses to form a list of protein with high probability of correct identification.

### 1.3.4 Surface enhanced desorption/ionization mass spectrometry

SELDI which employed chromatography and mass spectrometry was introduced by Hutchens and Yip in 1993 (Hutchens and Yip 1993). It has emerged as one of the most popular mass spectrometry based techniques for biomarker discovery because of its ease of use, high throughput, automation and the capability to analyze multiple samples in small volumes in a short time. Basically, it consists of three major components: the ProteinChip array, the ProteinChip reader and the ProteinChip software. The analysis of a sample using SELDI is aimed to reduce the sample complexity by selectively capturing proteins by different active surfaces of an array. There are two main types of active surfaces, which are the chemical surface and pre-activated biochemical surface.

In general, a protein sample is added to the protein array and incubated for a certain period. Only a subset of proteins in the sample binds to the protein chip while the unbound proteins are washed away. A matrix is added to the chip to co-crystallized with the bound proteins and subsequently proceed to the ProteinChip Reader.

The ProteinChip Reader is a laser desorption ionization time of flight mass spectrometry instrument equipped with a pulsed ultraviolet nitrogen laser source that uses state-of-the-art ion optic and laser optic technology. The ionization and desorption processes that occur in the ProteinChip Reader are similar to the MALDI-TOF. Eventually, the detected proteins are presented as a series of peak with *m/z* values and intensities.

The ProteinChip Software is used to control the functions of the ProteinChip Reader and facilitates data collection and processing. This includes the loading of the ProteinChip array into the equipment, calibrating the reader, analyzing the samples, normalizing spectra and presenting the data.

## 1.4 Glycosylation

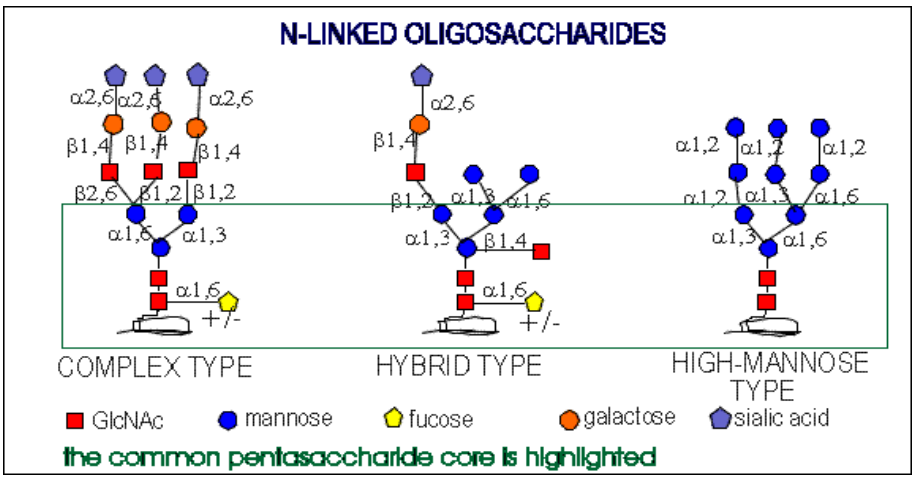
### 1.4.1 Biosynthesis and structure

Glycosylation is an extensive post-translational modification that can alter the function of proteins and thus the characteristics of the cell. There are two main classes of glycosylation which are *N*- and *O*-glycosylation. *N*-glycosylation involves the binding of *N*-acetylglucosamine to the amine side chain of asparagines, with sequence Asn-X-Ser(Thr), where X can be any amino acid except proline. On the other hand, *O*-glycosylation involves the binding of *N*-acetylgalactosamine to the hydroxyl of serine or threonine residue. Glycosylation is protein core specific, species specific and tissue specific. Thus, its activity is mainly controlled by a group of enzymes called glycosyltransferase. The glycosyltransferases are classified based on their substrates and the basis of the structures they recognize as an acceptor. The glycosyltransferases

catalyze the transfer of a sugar residue from a nucleotide to the particular biomolecules (protein or lipid) (Peracaula *et al.*, 2008).

*O*- and *N*-glycosylation occur through several discrete steps that incorporate sugars using a glycosyltransferase. For *N*-glycosylation, it is initiated by an oligosaccharide that comprise two *N*-acetylglucosamine, nine mannose and three glucose residues in the rough endoplasmic reticulum. This 14 carbohydrate residues precursor is carried by a dolichol pyrophosphate, followed by the transfer of the oligosaccharide to an asparagine residue of the polypeptide chain. Three glucose residues and one mannose residue are then removed by a specific glycosidase and consequently gives rise to three major classes of *N*-linked oligosaccharide: high mannose oligosaccharide, hybrid oligosaccharide and complex oligosaccharide by several enzymatic sequential pathways (Figure 1.1) (Ghazarian *et al.*, 2011). The high mannose oligosaccharide contains branches with two to six additional mannoses linked to the pentasaccharide core. The hybrid oligosaccride consists of one complex branch and one or more high mannose branches while the complex oligosaccharide contains two or more branches with each of them consisting of at least one *N*-acetylgalactosamine, one galactose and one sialic acid to form bi-, tri- or tetraantennary structures (Durand and Seta 2000).

**Figure 1.1: Three different structures of *N*-linked oligosaccharides**




* Image was reproduced with permission from www.crystbbk.ac.uk

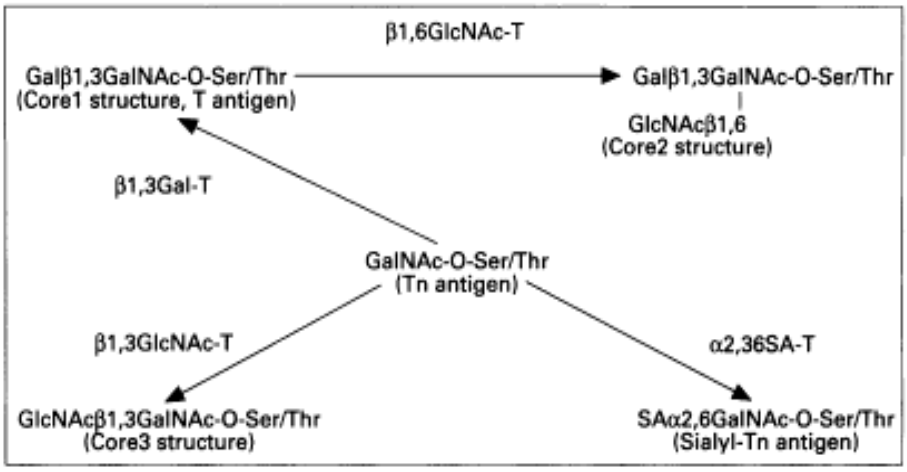
In case of *O*-glycosylation, *N*-acetylgalactosamine is covalently linked to the serine or threonine of a polypeptide chain. This structure forms the Tn antigen if further elongation does not occur. However, the elongation with galactose linked to the O-3 position creates the core 1 structure, which is also known as the Thomsen-Friedenreich epitope. The core 2 and 3 structures are formed with the covalently β1,6 or β1,3-linked of N-acetylglucosamine to N-acetylgalactosamine (Figure 1.2) (Dall'olio 1996). These core structures can be further elongated by the addition of other carbohydrates such as galactose, fucose and N-acetylglucosamine to form a stable *O*-linked glycan (Durand and Seta, 2000). The sialyl-Tn antigen is created by addition of sialic acid at the O-6 position of the N-acetylgalactosamine residue (Dall'olio 1996).

### 1.4.2 Glycosylation and its correlation to cancers

Glycosylation plays many important roles in order to control the stability and the function of the cell. It helps to maintain the homeostasis of the cell by being involved in processes such as being cell cycle regulators, anti-apoptotic proteins, transcription factors and tumor suppressors. Furthermore, it stabilizes proteins for their proper folding, orientation and conformation. Lastly, glycosylation also functions as signaling molecules, recognition molecules and adhesion molecules (Hakomori, 1985).

Because of its enormous biological functions, aberrant glycosylation of glycoproteins is one of the most frequent changes that occurred in the cancer biological system. The importance of the phenomenon is to provide better survival condition for cancer cell invasion. Changes of glycosylation are usually related to three main types of patterns including, the synthesis of highly branched and heavily sialylated glycans, the expression of incomplete forms and the re-expression of glycosidic antigens of the fetal

**Figure 1.2: Common core structures of *O*-linked oligosaccarides**




Abbreviations:

GalNAc=N-acetylgalactosamine;

β1,3Gal-T=β1,3-galactosyltransferase;

β1,6GlcNAc-T=β1,6-N-acetylglucosaminyltransferase;

β1,3GlcNAc-T=β1,3-N-acetylglucosaminyltransferase;

α2,6SA-T=α2,6-sialytransferase

Image were reproduced with permission from Dall'olio 1996.

type (Dall'olio, 1996). Aberrant glycosylation is more likely caused by the altered regulation of one or more key glycosyltransferase in the biosysnthesis of glycoproteins. For example, the GnT-III and GnT-V enzymes have their increased of activity in the formation of β1,6-branched Asp-linked oligosaccharides in hepatocellular carcinoma (Dennis *et al.,* 1987; Fernandes *et al.,* 1991). This structure has been previously reported to correlate with the cancer progression. Another example is the overexpression of GlcNAc-TV enzyme for the biosynthesis of β1,6-GlcNAc-branched, which has been shown to induce the metastatic potential of carcinoma cells of patients with breast cancer (Dennis *et al.,* 1999).

## 1.5 Gynaecological cancer

Cancer is a group of cells that divides uncontrollably due to deactivation of programmatic death process and has the capability to spread to another part of the body through the blood system. It becomes the most prevalent disease in the world and the leading cause of death. According to the prediction of the International Agency for Research on Cancer (IARC), 27 million newly diagnosed cancer cases and 17 million cancer deaths will occur each year by 2030. In fact, statistics of the worldwide cancer incident is frightening as a total of 12.7 million new cancer cases and 7.6 million cancer deaths have been reported in 2008 (Ferlay *et al.,* 2010).

Gynaecological cancer refers to those malignant cancer cells that affect the organs and originated from the female reproductive organ, which include endometrium, cervix, ovary, vagina and vulvar. Figure 1.3 shows the female reproductive organ. Endometrial, cervical and ovarian cancers are listed in the ten most common cancers affecting women worldwide (Table 1.2).

**Figure 1.3: Female reproductive organ**

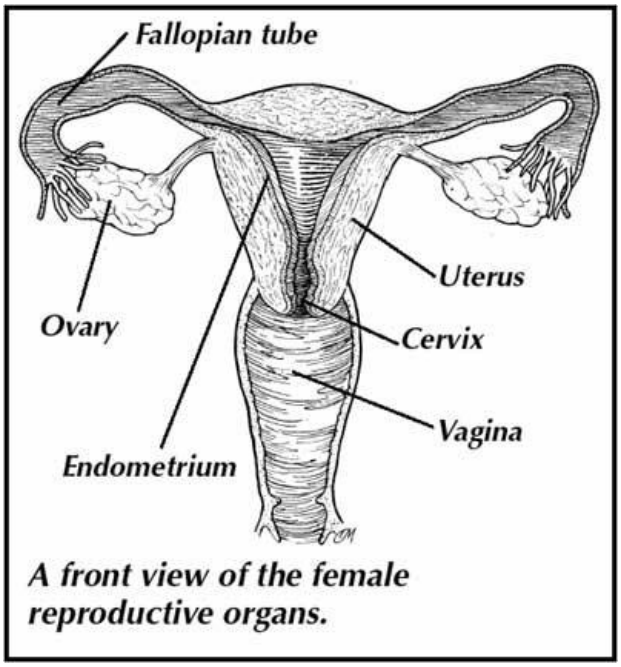



***Image was reproduced with permission from http://reproductivesysteminfo.com**

**Table 1.2: World ten most common cancer in women (Ferlay *et al.*, 2010)**

| Type of cancer | Number ('000) | Percentage (%) | ASR (per100,000) |
|---|---|---|---|
| Breast | 1383 | 22.9 | 39.0 |
| Colon | 570 | 9.4 | 14.6 |
| Cervix | 529 | 8.8 | 15.2 |
| Lung | 513 | 8.5 | 13.5 |
| Stomach | 349 | 5.8 | 9.1 |
| Endometrial | 287 | 4.8 | 8.2 |
| Ovary | 225 | 3.7 | 6.3 |
| Liver | 225 | 3.7 | 6.0 |
| Thyroid | 163 | 2.7 | 4.7 |
| Non-Hodgkin lymphoma | 156 | 2.6 | 4.2 |

ASR: Age standardized indices

### 1.5.1 Endometrial cancer (ECa)

#### 1.5.1.1 Epidemiology

According to IARC GLOBOCAN 2008, endometrial cancer (ECa) is the 6$^{th}$ most common cancer among women worldwide. There were a total of 287,000 new cases that had been reported in 2008, which constituted 4.8% of the total case of women cancer. The age standardized indices (Zareian, Khajeh et al.) were 8.2 per 100,000 population for females, with a total of 74,000 death cases reported. The incidence of ECa is highest in the developed region, with 50% of the total cases reported from this

region. ECa has afavourable prognosis with a five year survival rate of approximately 80-90% in the developed countries and 70% in the developing countries.

In the year 2006, ECa was the 7$^{th}$ most frequent cancer among women in Peninsular Malaysia (Omar *et al.,* 2006). It constituted 3.2% of the total female cancer cases which was equivalent to 372 cases registered under the Malaysian Cancer Statistics. The ASR was 4.4 per 100,000 population for female. The incidence rate for ECa was highest among the Malaysian Chinese population (5.2 per 100,000 population), followed by Malaysian Indians (4.7 per 100,000 population) and the Malays (2.8 per 100,000 population). Its incidence rate increased with age after 30 years old and achieved its peak at the age groups of 60-69 years old.

#### 1.5.1.2 Pathological classification

ECa originates from glandular tissue within the inner lining of the uterus. It is a heterogenous disease and therefore can be distinguished into two biologically different subtypes based on their clinical features and pathological appearance. Type I ECa is associated with endometrial hyperplasia and grouped as endometrioid carcinomas. It is the most common type of endometrial cancer which develops from endometrioid cells that form glands in the lining of the uterus (Sherman *et al.,* 1995). Type II ECa develops from a surface lesion and is a type of endometrial intraepithelial carcinoma called as serous carcinoma. It is poorly differentiated, occurs spontaneously in postmenopausal women and more aggressive in tumour invasion (Purdie, 2003).

### 1.5.1.3 Etiology

The precise cause of ECa remains unknown, even though several risk factors have been determined. ECa is known to be related to an increase in the levels of estrogen and obesity (Kaaks *et al.,* 2002). Therefore, women who have never been pregnant, undergone early puberty or late menopause and exposed to high level of estrogen are more likely to contract ECa. Exposure to tamoxifen has also been reported to be associated with high rates of ECa. Tamoxifen is a drug which has properties similar to estrogen and is prescribed for the prevention or treatment of breast cancer. In addition, a diet with a high intake of animal fats also increases the risk of ECa. Women with a history of breast, ovarian and colon cancer or have a family history of ECa are also at high risk (Schoen *et al.,* 1994). Lynch Syndrome, which is also known as hereditary nonpolyposis colorectal cancer syndrome, can increase the risk through the inheritance of genetic mutation associated with ovarian and colon cancer (Bewtra *et al.,* 1992). However, the correlation of breast cancer and ECa remained obscure.

### 1.5.1.4 Staging

The surgical staging of ECa was first introduced by the Federation of International Gynecologists and Obstetricians (FIGO) in 1988. Recently, the FIGO staging system for endometrial cancer was revised as shown in Table 1.3 (Creasman, 2009).

**Table 1.3: FIGO Staging of ECa (Creasman, 2009)**

| Stage | Description |
|---|---|
| IA | Tumor confined to uterus, <50% myometrial invasion |
| IB | Tumor confined to uterus, ⩾50% myometrial invasion |
| II | Cervical stromal invasion |
| IIIA | Tumor invasion into serosa or adnexa |
| IIIB | Vaginal or parametrial involvement |
| IIIC1 | Pelvic node involvement |
| IIIC2 | Paraaortic node involvement |
| IVA | Tumor invasion into bladder or bowel mucosa |
| IVB | Distant metastases (including abdominal metastases) or inguinal lymph node involvement |

#### 1.5.1.5 Screening and diagnosis

Currently, the diagnosis of ECa is very dependent on endometrial biopsy, which involves suction of the uterine wall tissue for endometrial histology evaluation. However, it is invasive, highly subjective and not adequately sensitive and specific for the routine diagnosis of ECa. Another method for the diagnosis of ECa is transvaginal ultrasound. Through the ultrasound imaging, the thickness and abnormalities of the endometrial cells can be detected with good sensitivity (Jacobs *et al.*, 2011). The most advanced detection method for ECa is magnetic resonance imaging, which can accurately assess the histology staging of the cancer and tumor response to therapy and differentiate tumor recurrence from post-treatment changes (Beddy *et al.*, 2012). Up until now, no tumor marker has been established specifically for ECa, even though several serum markers such as CA125 , CA19-9, CA15-3, CA72-4 and CEA that are

known to be non-specific are currently being applied for its diagnosis (Ueda *et al.*, 2010).

### 1.5.2 Ovarian Cancer (OCa)

#### 1.5.2.1 Epidemiology

Ovarian cancer (OCa) is the 7$^{th}$ most common cancer in women, with a total of 225000 new cases reported in 2008 (Ferlay *et al.*, 2010). Although OCa is the third most common gynaecolgical cancer after ECa and CCa, it is the most fatal malignancy due to its late presentation of symptoms, weak response to treatment and high recurrence rate. Approximately 140,000 deaths or 4.2% of all women cancer death rates are due to this cancer. The incidence rate of OCa is higher in the developed region compared to the developing region. The ASR for developed and developing regions were 9.4 and 5.0 per 100,000 populations, respectively.

OCa is the fourth most common cancer for female in Malaysia (Lim and Yahya, 2003). In the year 2007, there were 658 women in the Peninsular Malaysia who were diagnosed with OCa. The incidence of OCa in Peninsular Malaysia increased with age with the ASR of 7.6 per 100,000 population. Among the three major races of Malaysians in the peninsular, the Malay women had the highest incidence rate for OCa with the ASR of 6.8 per 100,000 population. This is followed by the Malaysian Indians (6.4 per 100,000 population) and Chinese (5.8 per 100,000 population).

### 1.5.2.2 Pathological Classification

OCa is also a heterogenous malignancy that develops from diverse types of cells within the ovary. According to FIGO, OCa is classified into three main types, which are the epithelial tumour, the germ cell tumour and the sex cord stroma tumour (Brown, 2001). Epithelial tumour arises from the germinal epithelium lining of the ovary. It constitutes approximately 85% of all malignant ovarian tumours. It is further subdivided into several subtypes, including serous, mucinous, endometroid, clear cell, transitional cell, mixed and undifferentiated carcinoma. Around 10-20% of the epithelial tumours have low malignant potential, which is called borderline ovarian carcinoma. It is a histologic variant that is less aggressive without stromal invasion but only demonstrates papillary proliferation and multilayering of epithelial cells with nuclear abnormalities (Soslow, 2008).

Germ cell tumour accounts for 5% of all OCa and develops in the embryo or placenta structures. It usually occurs in women younger than 30 years old. The most common type of germ cell tumour is called dysgerminoma, which consists of glycogen-rich cells that resemble germinal epithelium in a stroma with lymphoid cells (Young *et al.,* 1994). Other less common histologic subtypes include endodermal sinus tumor, immature teratoma, choriocarcinoma and embryonal carcinoma.

Sex cord stromal tumour constitutes 5% of all OCa. It mainly develops in the connective tissues and supporting ovarian stroma. This tumour is generally less aggressive and usually secretes steroid hormones, such as estrogen and androgen (Reed *et al.,* 2010).

**1.5.2.3 Etiology**

One of the main causes of OCa is the accumulation of genetic mutations. About 10% of OCa cases are predominantly caused by mutations in the BRCA1 and BRCA2 genes which is known as breast-ovarian cancer syndrome (Gevensleben *et al.*, 2010). Both the BRCA1 and BRCA2 are tumour suppressor genes, which are important in the regulation of the cell growth and proliferation. Therefore, the loss of functions for both genes have led to further mutations and genomic instability followed by the initiation of carcinogenesis (Pradhan *et al.*, 2010). The primary function of the BRCA genes is to preserve the structural and numerical stability of chromosomes during cell division. The BRCA1 expression proteins are responsible for repair of the double strand DNA breaks with error free homologus recombination. On the other hand, BRCA2 functions as mediator of interaction for homologus recombination.

The second type of inherited genetic mutations associated with OCa is the hereditary nonpolyposis colorectal cancer or Lynch syndrome. Patients with this syndrome are at an increase risk with cancers of gastrointestinal tract organs, upper urological tract and female genital organs including the ovary and endometrium (Watson and Lynch, 1993). Lynch syndrome is associated with mutations in the mismatch repair genes located within the DNA mismatch repair pathway (Lu, 2008). This eventually affects the stability of the genome and the function of the tumour suppressor genes.

The development of ovarian malignancy which also refers as sporadic ovarian cancer has been identified to be correlated to genetic alteration in a complex pathway involving multiple oncogenes and tumour suppressor genes, including p53, p16, p27, HER2/*neu*, *fms*, *ras*, *c-myc*, *myb*, *BRAF*, *KRAS*, *PTEN* and M-CSF (Gallion *et al.*,

1995). They are altered in expression or in their various mechanisms such as gene amplification, point mutation and chromosomal translocation (Kacinski *et al.*, 1990; Gallion *et al.*, 1995; Shih Ie and Kurman, 2004).

Besides genetic factors, both the endogenous factors, including age, family history with ovarian, breast, cervical or colon cancer, infertility, early puberty, late menopause and hormone, as well as the exogenous factors, including environmental factors, diet, lifestyle and biological and chemical exposure have also been reported to increase the risk of OCa (Booth *et al.*, 1989; Daly and Obrams, 1998). However, the risk of OCa is reduced when a woman has her pregnancy at an early age, early menopause and used oral contraceptives (Booth *et al.*, 1989).

**1.5.2.4 Staging**

The current staging for OCa is the FIGO system developed in 1987. It is based on evaluations obtained from the surgical exploration of ovarian tissues and its surrounding. Table 1.4 shows the FIGO staging of OCa.

**Table 1.4: FIGO staging of OCa (FIGO, 1987)**

| Stage | Description |
|---|---|
| I | Growth limited to ovaries only |
| IA | One ovary involved with intact capsule and no tumor on external surface |
| IB | Both ovaries involved with intact capsule and no tumor on external surface |
| IC | One or both ovaries with tumor on surface and capsule ruptured |
| II | Pelvic extension |
| IIA | Metastasis to uterus and/or fallopian tubes |
| IIB | Metastasis to other pelvic tissues |
| IIC | Metastasis to uterus, fallopian and pelvic tissues with tumor on surface or capsule ruptured |
| III | Tumor involved one or both ovaries with peritoneal implants outside the pelvis and/or positive retroperitoneal or inguinal nodes |
| IIIA | Tumor limited to the true pelvis with negative nodes but histologically confirmed microscopic seeding of abdominal peritoneal surfaces |
| IIIB | Tumor involved one or both ovaries, histologically confirmed implants on abdominal peritoneal surfaces <2cm |
| IIIC | Abdominal implants >2cm and/or positive retroperitoneal or inguinal nodes |
| IV | Growth involved one or both ovaries with distant metastasis |

### 1.5.2.5 Screening and diagnosis

Currently, pelvic examination, transvaginal ultrasound screening of the ovary and tumor marker panels such as CA125 detection are used to screen and diagnose OCa. The application of pelvic examination is still not well established due to its low sensitivity and specificity for detection of the cancer. Transvaginal ultrasound with or without power Doppler is used to identify abnormalities of the ovary and is capable to

distinguish benign ovary and malignant masses of OCa. It is highly sensitive but is not cost effective and has an unsatisfactory specificity (Menon, 2004). CA125 is recognized by FIGO as one of the approved tumour markers for OCa. The level of CA125 is elevated for approximately 90% of patients with advanced OCa but only 50% of patients with early stage of OCa (Gupta and Lis, 2009). In addition, the CA125 has limited specificity in detecting OCa as it is also raised in other types of cancers and gynecologic or non-gynaecologic conditions as well. Although the combination of CA125 and transvaginal ultrasound can improve the sensitivity and specificity for early detection of OCa, many cases of OCa are only detected at the advance stages of the cancer, which are associated with high rates of mortality.

### 1.5.3 Cervical Cancer (CCa)

#### 1.5.3.1 Epidemiology

CCa is the second most frequent cancer affecting women worldwide (Ferlay *et al.*, 2010). In 2008, approximately 529,000 women were newly diagnosed with CCa, which constituted roughly 8.8% of all female cancers. The mortality rate was 0.52 with a total of 275,000 patients died caused by CCa. Indeed, it is considered as a leading cause of death for women in the developing countries, which accounts for about 88% of cancer death that occurred in the region. The incidence rate of CCa was the highest in the developing region, particularly Africa, South America and South Central Asia.

In Malaysia, CCa is ranked the third most common cancer after breast cancer and colorectal cancer among women. But in Peninsular Malaysia, it was the fourth most frequent cancers, constituting 4.9% of the total cancer cases registered with the National Cancer Registry in the year 2006 (Omar *et al.*, 2006). Current estimates

indicate that approximately 1,492 Malaysian women are detected with CCa every year. More than half of the women were diagnosed at an advanced stage and 766 die from the disease. The Social Statistic Bulletin of Malaysia 2005 reported that the death rate for CCa was rising from 0.29% in 1996 to 0.41% in 2000 (Abdullah and Su, 2010). The overall ASR was 12.2 per 100,000 population. CCa commonly affected females with ages more than 30 years and had its peak at the age group of 60-69 years. The Malaysian Chinese women had the highest ASR, which was 13.6 per 100,000 population. This was followed by the Malaysian Indians (11.3 per 100,000 population) and the Malays (7.3 per 100,000 population).

#### 1.5.3.2 Pathological classification

CCa originates at the squamocolumnar junction in the cervix. It can be differentiated into two groups, which are squamous cell carcinoma and adenocarcinoma (Kaur *et al.,* 2003). Squamous cell carcinoma is the most common subtype of CCa, which accounts 85% of the total CCa's incidence. It arises from the surface lining of the ectocervix connected to the endocervix. On the other hand, adenocarcinoma arises within glands located in the endocervix.

#### 1.5.3.3 Etiology

Studies have shown that the human papillomavirus (HPV) infection is the main causative agent for more than 90% cases of invasive CCa worldwide (Franco *et al.,* 2001). The HPV infection affects women who are sexually active and have weakened immune system. HPV can be divided into two distinct categories, which are the low risk and high risk types. The low risk type rarely cause development of CCa. On the other hand, the high risk HPV including types 16, 18, 31 and 45 give rise to the high

grade cervical intraepithelial neoplasia, which is a precursor of invasive CCa. HPV-induced cancers express two viral oncoproteins E6 and E7, which interact with p53 and the tumor suppressor gene product pRb, respectively. This ultimately induces genetic instability and thus leads to carcinogenesis of the cervical cells by activation and inactivation of the tumour suppressor genes (Senba and Mori, 2012).

#### 1.5.3.4 Staging

The classification of CCa is mainly based on clinical evaluation of the cervix lesions. Table 1.5 shows the revised FIGO staging system for CCa that is currently being used.

#### 1.5.3.5 Screening and diagnosis

Screening of CCa is currently performed by the Papinicolaou (Pap) smear method. Pap smear, which was introduced in Malaysia 30 years ago, detects cytologic abnormalities of lesions collected from the end of the uterus extending into the vagina. However, the popularity of this method is still unsatisfactory among Malaysian women.

According to a report from the National Health and Morbidity Survey 1996, only 26% of Malaysian women have been covered by this method (Wong *et al.,* 2009). To date, it is still the most important and useful tool for prevention of CCa. Yet, it is highly subjective and has low sensitivity and specificity (Boulet *et al.,* 2008).

**Table 1.5: FIGO staging of CCa (Odicino *et al.*, 2008)**

| Stage | Description |
|---|---|
| IA | Invasive carcinoma diagnosed only by microscopy. All microscopically visible lesions even with superficial invasion are stage IB. |
| IA1 | Stromal invasion no greater than 3.0 mm in dept and 7.0 mm or less in horizontal spread. |
| IA2 | Stromal invasion greater than 3.0 mm and not more than 5.0 mm with a horizontal spread of 7.0 mm or less. |
| IB | Clinically visible lesion confined to the cervix or microscopic lesion greater than IA2. |
| IB1 | Clinically visible lesion 4.0 cm or less in greater dimensions. |
| IB2 | Clinically visible lesion more than 4.0 cm in dimension. |
| II | Tumor invades beyond the uterus but not to the pelvic wall or to lower third of the vagina. |
| IIA | Without parametrial invasion. |
| IIB | With parametrial invasion. |
| III | Tumor extends to pelvic wall and/or involves lower third of vaginal or causes hydronephrosis or non-functioning kidney. |
| IIIA | Tumor involves lower third of vagina, no extension to pelvic wall. |
| IIIB | Tumor extends to pelvic wall and/or causes hydronephrosis or non-functioning kidney. |
| IVA | Tumor invades mucosa of bladder or rectum and/or extends beyond the true pelvic. |
| IVB | Distant metastasis |

As mentioned earlier, HPV is responsible for the carcinogenesis of CCa. Therefore, it has been suggested that HPV testing can improve the efficiency of CCa screening. Currently, there are two types of detection technologies for HPV screening, which are Hybrid capture 2 and PreTect HPV-Proofer (Keegan *et al.,* 2009). Hybrid capture 2 is a signal amplified hybridization microplate-based assay used to detect 18 types of high and low risk HPV types using probe cocktails (Poljak *et al.*, 2002). It was

recognized and approved by FDA in 2003 as a primary screening test for CCa as it provides sufficient scientific evidences to support its performance in clinical diagnostics. On the other hand, PreTect HPV-Proofer is a real-time based multiplex NASBA used to detect the HPV type specific E6/E7 mRNA from the HPV high risk types 16, 18, 31, 33 and 45 (Molden *et al.,* 2007). Compared with the Hybrid capture 2, it is more specific in identifying women with high grade cervical intraepithelial neoplasia 2+ but has a lower sensitivity (Ratnam *et al.,* 2010). Recently, a novel method known as MassARRAY spectrometry utilizes competitive PCR, primer extension and MALDI-TOF/TOF to detect 15 types of high risk HPV (Basu *et al.,* 2011). However, this method is still under assessment for clinical application.

Another approach for the diagnosis of CCa is the utilization of serum biomarkers. The squamous cell carcinoma antigen is the most commonly used serum marker for squamous cell cervical carcinoma, followed by the serum fragment of cytokeratin 19 (CYFRA 21-1). Pre-treatment serum levels of both markers have been demonstrated to be correlated with the disease stage, tumor size, depth of the stromal invasion, the lymph-vascular space involvement and lymph node metastasis (Ueda *et al.,* 2010). Similarly, carcinoembryonic antigen (CA), CA 125, CA 19-9 have been used to detect the adenocarcinoma even though they have inadequate positive rates toward to the disease (Ueda *et al.,* 2010). In general, the use of specific markers for CCa in conjunction with Pap smear and HPV testing can improve the accuracy, precision and sensitivity of the screening program for CCa.

## 1.6 Problem statement and rationale of the study

The diagnosis of ECa, OCa and CCa currently relies on existing tumor markers that are not adequately sensitive and specific for early detection of the cancers. In view of these limitations, there is a need to identify new biomarkers that are more specific and sensitive to assist in the early and accurate identification of the cancer patients, which may improve the rates of mortality of the cancers. Urine of the patients may contain aberrant proteins that can be used as new biomarkers. Urinary biomarkers, if identified, provide the advantage of performing the diagnosis non-invasively.

## 1.7 Research objectives

The aim of the present study was to analyze urinary proteins from patients with gynecological cancers, including ECa, OCa and CCa, for abnormally glycosylated or excreted glycoproteins using the champedak lectins as probes. The aberrantly glycosylated or expressed urinary proteins that were recognized by the lectins are potential biomarkers that may be used in the screening detection and/or monitoring of the cancers.

The specific objectives of the study were:

1) To develop 2-DE profiles of urinary proteins from control and patients with ECa.
2) To compare 2-DE profiles of urinary proteins from control and patients with ECa using Image Master 2D Platinum software.
3) To generate proteome maps of urinary *N*- and *O*-glycosylated proteins using western blotting cum lectin approach, lectin affinity chromatography, LC-MS/MS analysis and SELDI-TOF profiling.

4) To compare the glycoprotein maps of healthy control individuals with those generated from women with ECa, OCa and CCa.
5) To identify the abnormally gylcosylated or expressed glycoproteins that have the potential to be used as cancer biomarkers.

# Chapter 2
# Materials and Methods

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 Materials

### 2.1.1 Urine sample

Approximately fifty ml of morning midstream urine samples were obtained from patients who were newly confirmed with stage IB and IIA/B gynaecological cancers, ECa (n=7), OCa (n=9) and CCa (n=8), prior to treatment or surgery in the morning at the Gynaecology Clinic, University of Malaya Medical Centre, Kuala Lumpur. All patients were confirmed with negative diagnosis for other diseases and showed normal serum creatinine values. Control urine samples (n=11) were obtained from normal healthy volunteers with the same range of age (30 years old to 75 years old). Samples were collected with subjects' consents and approval (MEC. REF.NO.: 435.18) granted by the Ethical Committee (Institutional Review Board) of the Medical Centre and the Medical Research and Ethics Committee of the Ministry of Health, Malaysia, in accordance to the ICH GCP guideline and the Declaration of Helsinki. The urine samples were preserved by adding sodium azide to a final concentration of 20 mM prior to centrifugation at 10000 × *g* for 15 minutes at 4$^{o}$C. The samples were dialyzed against four changes of 2 L of distilled water for 12 hours at 4$^{o}$C before being aliquoted, freeze-dried and stored at -80$^{o}$C.

### 2.1.2 Champedak (*Artocarpus integer*)

Two champedak fruits that were used to isolate one batch of CGB and CMB lectins were obtained from a fruit orchard in Kota Bahru, Kelantan.

### 2.1.3 General materials

The materials used during the course of this project and their respective brands and suppliers are as listed. All chemicals were of analytical grade and were purchased from Sigma Aldrich Company, St. Louis, U.S.A. unless stated.

**Enzymes and substrates**

1. **Roche Molecular Biochemicals, Mannheim, Germany**

   Activated alkaline phosphatase (AP)

   NBT/BCIP tablets

2. **Promega, Madison, USA**

   Trypsin Gold, mass spectrometry grade

**Chromatographic materials**

1. **GE Healthcare Biosciences, Uppsala, Sweden**

   Preactivated CNBr Sepharose 4B

**Two-dimensional electrophoresis**

1. **GE Healthcare Biosciences, Uppsala, Sweden**

   IEF immobiline dry strips 11 cm (pH 3-10)

   IPG buffer (pH 3-10)

   Drystrip cover fluid

2. **BioRad Laboratories Inc., U.S.A.**

   Pre-stained broad range SDS-PAGE standard

**SELDI Biochemical Arrays and matrices**

1. **BioRad Laboratories Inc., U.S.A.**

   ProteinChip® array PS10

   α-cyano-4-hydroxycinnamic acid (CHCA)

**Commercial kits**

1. **Pierce, Rockford Illinois, U.S.A.**

   BSA Protein assay kit

   Metal enhanced DAB solution kit

### 2.1.4 Standard solutions

**Phosphate-buffered saline (PBS)**

**(170 mM NaCl, 3.4 mM KCl, 10.0 mM $Na_2HPO_4$, 1.8 mM $KH_2HPO_4$, pH 7.2)**

| | |
|---|---|
| NaCl | 9.36 g |
| KCl | 0.25 g |
| $Na_2HPO_4$ | 0.25 g |
| $KH_2HPO_4$ | 1.42 g |

The solution was made up to 1 L with distilled water and pH was adjusted to pH 7.2. The solution was stored at room temperature.

**Tris Buffered Saline (TBS)**

**(100 mM Tris-HCl, 0.9% NaCl)**

| | |
|---|---|
| Tris | 12.11 g |
| NaCl | 9.00 g |

Tris was dissolved in 900 ml distilled water and the pH of the solution was adjusted to 7.5 by addition of concentrated HCl. NaCl was then added and the solution was made up to 1 L with distilled water.

## 2.2 Methods

In the present study, the urine samples of control and patients with ECa, OCa and CCa were analyzed using three approaches, which include 2-dimensional electrophoresis-lectin based Western blotting, lectin affinity chromatography-liquid chromatography and SELDI-TOF. Figure 2.1 shows the overall flow of methodology used in this study.

### 2.2.1 Purification of champedak lectins

#### 2.2.1.1 Extraction of crude lectin from champedak seeds

The seeds of champedak were peeled, cleaned with distilled water and dried at room temperature. The dried seeds were ground to powder with a blender. Forty grams

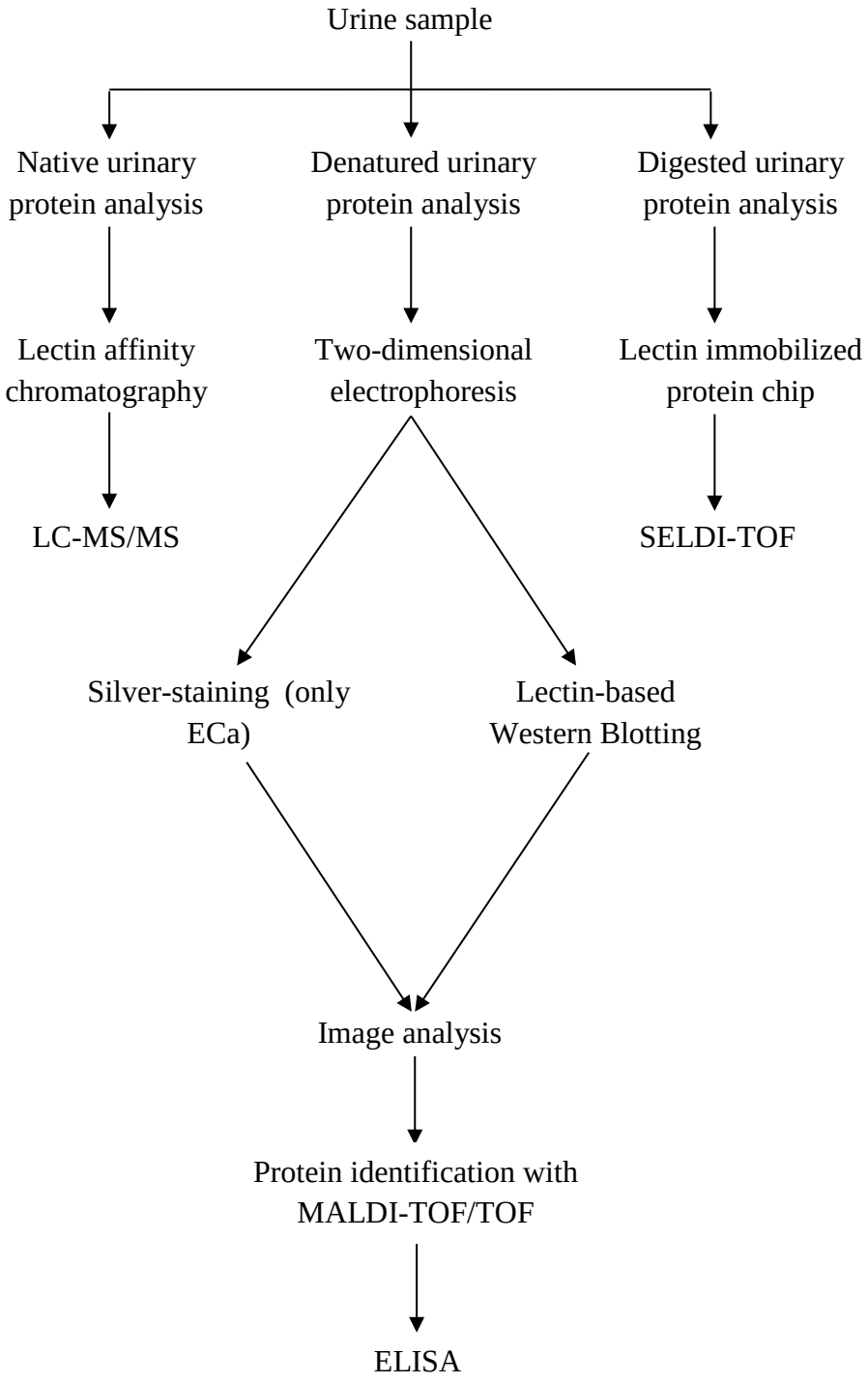



Figure 2.1: The overall flow of methodology adopted in the present study.

of the powdered seed were suspended in 400 ml of PBS pH 7.4 and stirred for 24 hours at 4°C. The homogenate was centrifuged at 8000 × *g* for 15 minutes at 4°C. Supernatant was collected and subjected to 60% ammonium sulphate precipitation and stirred for 2 hours at 4°C. The solution was then centrifuged at 8000 × *g* for 15 minutes at 4°C. The resulting pellet was collected and dissolved in cold PBS. Four times extensive dialysis of dissolved pellet was done in 48 hours. Finally, the crude extract was aliquoted and kept at -20 °C.

### 2.2.1.2 Preparation of Galactose and Mannose Sepharose-4B columns

**Activation of Sepharose-4B with Divinylsulfone**

Divinylsulfone (DVS) introduces reactive vinyl groups in Sepharose, which reacts with hydroxyl groups of sugar, causing covalently immobilization of sugars on Sepharose. A hundred ml of Sepharose-4B was washed with 1 L of distilled water in a sintered glass funnel, suctioned to a wet cake and transferred to a 500 ml beaker. The moist gel was suspended in 100 ml of 0.5 M sodium carbonate and stirred slowly in a fume cabinet. Ten ml of DVS was added slowly to the suspension over a period of 15 minutes in a fume cabinet with constant stirring. The gel was then stirred for 1 hour at room temperature before it was washed extensively with 1 L of distilled water until the filtrate was no longer acidic. The activated Sepharose was ready for ligand coupling.

**Coupling of sugar to DVS-activated Sepharose-4B**

Isolation of CGB lectin and CMB lectin were performed using immobilized galactose and mannose columns, respectively. Twenty ml of DVS activated gel was suspended in an equal volume of 20% (w/v) of D-galactose or D-mannose in 0.5 M

sodium carbonate. The mixture was stirred for 24 hours at room temperature. The gel was then washed with 2 L of water and 2 L of 0.5 M sodium bicarbonate. The gel was resuspended in 0.5 M sodium bicarbonate containing 2 ml of β-mercaptoethanol and stirred for 2 hours in a fume cabinet at room temperature. The gel was later washed with 2 L of distilled water, followed with 2 L of PBS. The resulting galactose-coupled and mannose-coupled Sepharose-4B gels were packed respectively into 2.8 cm diameter polypropylene columns of 5 cm in height.

### 2.2.2 Purification CGB and CMB lectins

#### 2.2.2.1 Isolation of CGB lectin

Prior to isolation of the CGB lectin, the galactose column was pre-equilibrated with PBS. Twenty ml of crude extract of champedak seeds was applied into the column and washed with PBS. Unbound fractions of 10 ml each were collected and absorbance was monitored at 280 nm. Peak fractions with absorbance more than 0.4 were pooled and kept for CMB lectin affinity isolation. The bound fractions containing CGB lectin were eluted using 1 ml of 0.8 M galactose in PBS, pH 7.4. The bound fractions with high absorbance more than 0.4 were pooled and dialysed against four times changes of PBS for 12 hours each. The dialysed bound fractions were freeze-dried, prior to storage at -20$^{o}$C.

#### 2.2.2.2 Isolation of CMB lectin

Isolation of the CMB lectin was performed using the unbound fraction collected from the previously mentioned CGB lectin isolation procedure. The fraction was loaded into a mannose column and loading and elution of the column was performed

using the same procedure that was described for the CGB lectin. A solution of 0.8 M D-mannose in PBS, pH 7.4 was used to elute the bound fraction containing the CMB-lectin. The dialysed bound fractions were freeze-dried, prior to storage at -20$^{o}$C. After elution of the desired lectins, both the galactose and mannose columns were regenerated by washing extensively with PBS before being reused.

#### 2.2.2.3 Preparation of cellulose membrane for dialysis

Cellulose membrane tubings were incubated in hot distilled water for three hours to remove glycerin, prior to use. The tubings were then treated with 0.3% (w/v) sodium sulfide to remove sulfur compounds at 80$^{o}$C for a minute and immediately washed with 60$^{o}$C distilled water for 2 minutes. The tubings were then acidified with 0.2% (v/v) of sulfuric acid for one minute and finally rinsed with hot water for another one minute before ready to be used.

### 2.2.3 Assessment of purified CGB and CMB lectins.

#### 2.2.3.1 Determination of the concentrations of CGB and CMB lectins.

Concentrations of the isolated lectins were estimated using a commercial protein assay kit based on the method of Bradford from Pierce, Illinois. The concentration estimation was performed according to the manufacturer's instruction. Bovine serum albumin (BSA) at 2 mg/ml was used as standard.

#### 2.2.3.2 Assessment of the purity of CGB and CMB lectins using SDS-PAGE

The purity of isolated lectins were evaluated by SDS-PAGE. The gel was stained with Coomassie Blue to obtain the protein profiles of the lectins.

**Standard solution for SDS-PAGE**

All solutions were stored at 4$^{o}$C, unless stated.

**Solution A: 30% acrylamide, 0.8% N,N-methylenebisacrylamide**

| | |
|---|---|
| Acrylamide | 60 g |
| N,N-methylenebisacrylamide | 1.8 g |
| Distilled water | topped up to 200 ml |

The solution was then stored in a dark bottle.

**Solution B: 1.5 M Tris-HCl, pH 8.8**

| | |
|---|---|
| Tris base | 36.23 g |
| Distilled water | 150 ml |

pH of the solution was adjusted to 8.8 using HCl and topped up to 200 ml.

**Solution C: 10% (w/v) SDS**

| | |
|---|---|
| SDS | 10 g |
| Distilled water | 100 ml |

The solution was stored at room temperature.

**Solution D: 10% (w/v) Ammonium persulphate (APS)**

| | |
|---|---|
| APS | 25 mg |

| | |
|---|---|
| Distilled water | 0.25 ml |

APS solution was prepared fresh prior to the preparation of the gel.

**Solution E: N, N, N', N' tetramethylethylenediamine (TEMED)**

**Solution F: 0.5 M Tris-HCl, pH6.8**

| | |
|---|---|
| Tris base | 6.1 g |
| Distilled water | 50 ml |

The solution was adjusted to pH 6.8 using HCl and topped up to 100 ml.

**4X SDS-PAGE sample buffer:**

**(62.5 mM Tris-HCl, pH 6.8, 10% (w/v) glycerol, 2% (w/v) SDS, 1% (w/v) DTT and 0.01% (w/v) bromophenol blue)**

| | |
|---|---|
| 0.5 M Tris-HCl, pH 6.8 | 2.5 ml |
| Glycerol | 2 ml |
| SDS | 400 mg |
| DTT | 200 mg |
| Bromophenol blue | 20 mg |

All chemicals listed above were mixed and topped up to 20 ml with deionized distilled water.

**Electrophoresis Buffer:**

**(25 mM Tris base, 198 mM glycine, 0.1% (w/v) SDS)**

| | |
|---|---|
| Tris base | 1.22 g |
| Glycine | 5.94 g |
| SDS | 0.4 g |

The solution was topped up to 1 L with distilled water.

**Fixing solution**

| | |
|---|---|
| 10% (v/v) acetic acid | 10 ml |
| 40% (v/v) methanol | 40 ml |

The solution was made up to 100 ml with distilled water.

**Hot Coomassie Blue stain**

| | |
|---|---|
| 10% (v/v) acetic acid | 10 ml |
| 0.1% (w/v) Coomassie Brilliant Blue (R250) | 0.1 g |

The chemicals were mixed and topped up to 100 ml. The staining solution was stored at room temperature and heated prior to use.

**Destaining solution**

10% (v/v) acetic acid solution

### 2.2.3.3 Preparation of stacking and separating gels

Glass plates were assembled and checked for leakage with distilled water before a mixture of 18% separating gel was poured into the sandwich. The mixture of separating gel was prepared according to the Table 2.1. The gel was overlaid with a layer of distilled water to achieve an even gel surface. Once the gel polymerized, the overlaid water was removed with a filter paper. The stacking gel mixture was prepared as stated in Table 2.1. It was layered on top of the separating gel and a gel comb was inserted into it before it was polymerized.

**Table 2.1: Volumes of stock solution used to prepare the separating and stacking gels of SDS-PAGE**

| **Stock Solution** | **Volumes of stock solutions for** | |
|---|---|---|
| | **Separating gel (18%)** | **Stacking gel (4%)** |
| Solution A (ml) | 12.00 | 0.65 |
| Solution B (ml) | 5.00 | - |
| Solution C (µl) | 200.00 | 50.00 |
| *Solution D (µl) | 100.00 | 25.00 |
| *Solution E (µl) | 6.60 | 5.00 |
| Solution F (ml) | - | 1.25 |
| Distilled water (ml) | 2.69 | 3.05 |

| | | |
|---|---|---|
| Total volume (ml) | 20.00 | 5.03 |

*The solutions were added prior to use

### 2.2.3.4 SDS polyacrylamide electrophoresis

Samples were mixed at a ratio of 3:1 with sample buffer in a microcentrifuge tube and boiled at 95$^{o}$C for 5 minutes. Ten µl of each samples and broad range molecular weight standard (Biorad Laboratories Inc., U.S.A.) were loaded into the wells. Electrophoresis was conducted at a constant voltage of 100 V and stopped once the blue front dye approached 1 cm from the gel bottom. The gel was then stained with hot Coomassie staining protocol.

### 2.2.3.5 Hot Coomassie staining protocol

The fixing solution was poured into a container containing the gel and fixed overnight on a shaker. The Coomassie stain was heated up to 90$^{o}$C in a microwave oven. The fixing solution was poured out and replaced with the staining solution followed with 30 minutes shaking. The staining solution was removed from the container and 20 ml of destaining solution was then added. The gel was left to shake overnight before being rinsed with distilled water and scanned.

## 2.2.4 Two-dimensional gel electrophoresis

Separation of the proteins using two-dimensional gel electrophoresis involves two steps. For the first dimension, the proteins were separated based on their isoelectric

points (p*I*) and in the second dimension, separation of the electrofocused proteins was based on their molecular weights.

### 2.2.4.1 First dimension electrophoresis

**Stock solution:**

**Rehydration buffer**

**(8 M urea, 2% NP40, 0.5% Pharmalyte, 0.002% Orange G)**

| | |
|---|---|
| Urea | 12 g |
| NP40 | 120 µl |
| Pharmalyte | 125 µl |
| 1% Orange G | 50 µl |

All chemicals were mixed, made up to 25 ml and stored in 2.5 ml aliquots at -20$^{o}$C.

### 2.2.4.2 Preparation of sample for rehydration

The concentration of the freeze dried sample was estimated based on the method as described in Section 2.2.3.1. A tube of rehydration solution was thawed and 7 mg of DTT was added into the solution. Three hundred µg protein was dissolved with 200 µl rehydration solution. Each sample was analysed once. The sample mixture was loaded into the slot of the reswelling tray. An 11 cm pH 3-10 linear IPG strip was put into the slot with the gel facing down on the sample mixture. Drystrip cover fluid was pipetted to cover the whole slot and the strip. The IPG drystrips were left to reswell for 16 hours at room temperature.

### 2.2.4.3 Isoelectric focusing

The first dimension separation was conducted using IPGphor (GE Healthcare Biosciences, Uppsala, Sweden). The temperature was set at 20$^{o}$C throughout the whole experiment. The rehydrated IPG strips were placed on the tray across the cathodic and anodic ends. Moistened electrode strips were placed on both ends of the rehydrated IPG strips. Both electrodes were put on both end of strip touching the moistened electrode strips. The IPG strips were covered with Drystrip cover fluid to ensure good thermal contact.

The first dimension isoelectricfocusing was conducted according to the recommended protocol by the manufacturer as shown below.

| Phase | Voltage | mA | W | V/hr |
|---|---|---|---|---|
| 1 | 100 | 50 | 5 | 100 |
| 2 | 1000 | 50 | 5 | 1000 |
| 3 | 8000 | 50 | 5 | 12500 |

Once the first dimension electrophoresis was completed, the focused strips were individually stored at -80$^{o}$C in screw cap tubes and ready for the second dimension electrophoresis.

### 2.2.4.4 Second dimension electrophoresis

**Stock solutions**

**SDS equilibration solution**

| | |
|---|---|
| 4X resolving gel buffer | 20 ml |
| Urea | 144.14 g |
| Glycerol | 138 ml |
| SDS | 8 g |

The solution was made up to 400 ml with distilled water.

**Electrophoresis buffer**

**(25 mM Tris-HCl, pH 8.3, 192 mM glycine, 0.1% SDS)**

| | |
|---|---|
| Tris base | 3.03 g |
| Glycine | 14.4 g |
| SDS | 1.0 g |

The solution was made up to 1L with distilled water.

**0.5% agarose sealing solution**

| | |
|---|---|
| Agarose | 0.5 g |
| Electrophoresis buffer | 100 ml |

The mixture was heated in a microwave oven until the agarose was dissolved completely.

#### 2.2.4.4.1 Preparation of 12.5% SDS PAGE gel

Glass plates were assembled as a sandwich in the gel caster according to the manufacturer's manual. A linear gel of 12.5% acrylamide was prepared according to the recipe in Table 2.2. Solutions for casting of the gel are described in Section 2.2.3.3. The gel was overlaid with distilled water and allowed to polymerize.

**Table 2.2: Volumes of stock solutions for 12.5% gel**

| | **Volumes of stock solutions** |
|---|---|
| | **Separating gel (12.5%)** |
| Solution A (ml) | 2.606 |
| Solution B (ml) | 1.563 |
| Solution C (µl) | 62.500 |
| *Solution D (µl) | 31.250 |
| *Solution E (µl) | 2.063 |
| Distilled water (ml) | 1.988 |
| Total volume (ml) | 20.00 |

*added prior to use

#### 2.2.4.4.2 Equilibration of IPG strips

A hundred mg of DTT was dissolved in 10 ml of SDS equilibration buffer. Five ml of the solution was pipetted into a tube containing an IPG strip. Tubes were shaken for 15 minutes. The DTT-SDS equilibration solution was then discarded. Two hundred fifty mg of iodoacetamide was dissolved in 10 ml of the same buffer and the solution was added and shaking was further carried out for another 15 minutes. The IPG strip was quickly rinsed with electrophoresis buffer after the equilibration solution was discarded. The strip was then placed on top of the gel and sealed with agarose solution.

#### 2.2.4.4.3 Second dimension separation

Second dimension electrophoresis was performed in two steps as shown below:

| Step | Voltage (V) | Current(mA) | Walt (W) |
|---|---|---|---|
| 1 | 50 | 17* | 2* |
| 2 | 600 | 25* | 15* |

*per gel

The whole experiment was conducted at a constant temperature of 16$^{o}$C. After electrophoresis, the gel was removed carefully from the cassette and was ready for staining or Western blotting.

#### 2.2.4.4.4 Silver staining of 2-DE gels

**Stock solutions**

**Fixing solution**

**(40% (v/v) ethanol, 10% acetic acid)**

| | |
|---|---|
| Ethanol | 400 ml |
| Acetic acid | 100 ml |

The solution was made up to 1 L with distilled water.

**Sensitizing solution**

**(30% (v/v) ethanol, 0.5 M sodium acetate, 8 mM sodium thiosulphate, 0.5% (v/v) glutaraldehyde)**

| | |
|---|---|
| Ethanol | 300 ml |
| Sodium acetate trihydrate | 68 g |
| Sodium thiosulphate | 20 g |

Solution was made up to 1 L of distilled water and 5 ml of glutaraldehyde was added prior to use.

**Silver solution**

**(14.25 mM silver nitrate, 0.04% (v/v) formaldehyde)**

| | |
|---|---|
| Silver nitrate | 2.5 g |

Solution was made up to 1 L with distilled water. Four hundred µl of formaldehyde was added prior to use.

**Developing solution**

**(0.4 M sodium carbonate, 5% sodium thiosulphate, 0.2% (v/v) formaldehyde)**

| | |
|---|---|
| Sodium carbonate | 25 g |
| Sodium thiosulphate | 28 µl |

Solution was made up to 1 L with distilled water. Four hundred µl of formaldehyde was added prior to use.

**Stopping solution**

**(40 mM EDTA-$Na_2$.2$H_2O$)**

| | |
|---|---|
| EDTA-$Na_2$.2$H_2O$ | 14.6 g |

Solution was made up to 1 L with distilled water.

**Preserving solution: 10% (v/v) glycerol**

**Staining protocol**

After completion of the 2-DE, gels were immersed in fixing solution for 30 minutes. The gels were then placed in sensitizing solutions overnight. After the incubation, the gels were washed three times for 5 minutes with distilled water. The silver reaction was carried out by immersing the gel in the silver solution for 40 minutes.

The gels were rinsed with distilled water to remove excess silver solution. The developing solution was then poured into the container which contained the gels. Once protein spots appeared, the gels were soaked in stopping solution for 30 minutes. The developed gels were washed with distilled water and stored in preserving solution at room temperature. For mass spectrometric analysis, the gels were stained in absence of glutaraldehyde.

### 2.2.5 Western blotting

Semi-dry transfer system (Novablot Kit for the Multiphor II Electrophoresis System, GE Healthcare Biosciences, Uppsala, Sweden) was used for the protein transfer process.

#### 2.2.5.1 Conjugation of CGB and CMB lectins to enzymes

To develop human urinary glycoprotein map, the urinary proteins which were resolved using 2-DE were electrophoretically transferred to the nitrocellulose membranes. The membranes were probed with enzyme conjugated lectins to form lectin-glycoprotein complexes. The complexes were then detected with colorimetric reagent specific for the enzymes.

##### 2.2.5.1.1 Conjugation of CGB lectin to Horseradish Peroxidase

**Stock solutions**

**0.1 M Sodium periodate**

| | |
|---|---|
| Sodium periodate | 21.39 mg |

The chemical was dissolved in 1 ml of deionised distilled water. The solution was prepared fresh prior to use.

**1 M Sodium acetate buffer, pH 4.4**

| | |
|---|---|
| Sodium acetate | 8.20 g/L |
| Acetic acid | 6.0 g/L |

Sodium acetate solution was mixed with acetic acid solution at a ratio of 1:2.

**0.1 M Sodium carbonate buffer, pH 9.5**

| | |
|---|---|
| Sodium carbonate | 10.6 g/L |
| Sodium hydrogen carbonate | 8.4 g/L |

Sodium carbonate solution was added to the sodium hydrogen carbonate solution until pH 9.5 was obtained.

**0.4% (w/v) Sodium borohydride**

| | |
|---|---|
| Sodium borohydride | 4.0 mg |

Sodium borohydride crystal was dissolved in 1 ml of distilled water.

**0.1 M borate buffer, pH 7.4.**

| | |
|---|---|
| Disodium tetraborate | 9.54 g/ 250 ml |
| Boric acid | 24.73 g/ 4L |

One hundred and fifteen ml of borate solution was added to 4 L of boric acid to pH 7.4.

**Conjugation protocol**

Four mg of horseradish peroxidase (HRP) was dissolved in 1 ml of distilled water. Two hundred µl of 0.1 M sodium periodate solution was added to the HRP solution and a green coloured solution was obtained. The solution was then dialysed overnight against 0.1 mM sodium acetate buffer, pH 4.4 at 4$^{o}$C. The dialysate was stirred for 20 minutes at room temperature and its pH was raised to 9.5 by the addition of 25 µl of 0.1 M sodium carbonate buffer, pH 9.5. This was immediately followed by the addition of 1 ml of CGB lectin (2 mg/ml) and the mixture was left to stir for 2 hours at 4$^{o}$C. The product was then reduced by addition of 100 µl of fresh sodium borohydride solution (4 mg/ml) and left to stand for another 2 hours at 4$^{o}$C. The HRP-conjugated CGB lectin was dialysed overnight against 0.1 M sodium borate buffer pH 7.4 at 4$^{o}$C. Finally, the conjugated lectin was diluted in equal volume of 60% glycerol in borate buffer and stored in eppendorf tube at 4$^{o}$C.

**2.2.5.1.2 Conjugation of CMB lectin to Alkaline Phosphatase**

**Stock solution**

**1 M sodium carbonate/hydrogen carbonate buffer, pH 9.4**

| | |
|---|---|
| Sodium carbonate | 10.6 g/100 ml |
| Sodium bicarbonate | 8.4 g/100 ml |

The pH of sodium bicarbonate was adjusted to 9.4 by the addition of sodium carbonate.

**100 mM sodium carbonate/hydrogen carbonate buffer, pH 9.8**

Ten ml of 1 M sodium carbonate/hydrogen carbonate buffer was diluted to 100 ml with deionised distilled water.

**200 mM sodium borohydride**

Eight mg of sodium borohydride was dissolved in 1 ml of cold deionised distilled water.

**2 M triethanolamine solution, pH 8.0**

Three ml of dionised distilled water was added to 2.66 ml of triethanolamine solution. The pH of the solution was adjusted to 8.0 with HCl and made up to 10 ml with deionised distilled water.

**1 M glycine, pH 7.0**

| | |
|---|---|
| Glycine | 0.75 g |

Glycine was dissolved in 6 ml deionised distilled water and the pH was adjusted to 7.0 with 0.1 M NaOH. The solution was then made up to 10 ml with deionised distilled water.

**Triethanolamine buffer, pH 7.6**

**(50 mM Triethanolamine, 150 mM NaCl, 1mM $MgCl_2$, 0.1 mM $ZnCl_2$, 10mM glycine and 0.1 % (w/v) sodium azide)**

| | |
|---|---|
| Triethanolamine | 16.65 ml |

| | |
|---|---|
| NaCl | 21.92 g |
| $MgCl_2$ | 0.51 g |
| $ZnCl_2$ | 34.0 mg |
| Glycine | 1.88 g |
| Sodium azide | 2.5 g |

The buffer was prepared by dissolving the above salts with 2.5 L of distilled water.

**Conjugation Protocol**

CMB lectin was concentrated to the final concentration of 7 mg/ml prior to the conjugation. Fifty µl of the concentrated CMB lectin was mixed with 100 µl of alkaline phosphatase (AP) solution and incubated overnight at 4$^{o}$C. Conjugation was terminated by adding 20 µl of 2 M triethanolamine solution, pH 8.0 followed by 40 µl sodium borohydride solution. The solution was stirred and incubated for 30 minutes at 4$^{o}$C. Five µl of triethanolamine was then added to the mixture and followed by 2 hours incubation at 4$^{o}$C. Ten µl of 1 M glycine solution pH 7.0 was added to the mixture. Then, the mixture was dialysed extensively against four changes of triethanolamine buffer. Sodium azide was added to the solution to the final concentration of 1 mg/ml. The alkaline phosphatase conjugated CMB lectin was aliquoted and stored at -80$^{o}$C.

**2.2.5.2 Protein transfer Western blotting**

**Stock Solution**

**Transfer buffer**

**(40 mM glycine, 0.1 M Tris, 0.038% (w/v) SDS, 20% (v/v) methanol)**

| | |
|---|---|
| Glycine | 2.93 g |
| Tris | 5.81 g |
| SDS | 0.375 g |
| Methanol | 200 ml |

The chemicals were dissolved and made up to 1 liter with distilled water.

**Tris buffered saline-Tween (TBS-T)**

**(TBS with 0.1% (v/v) Tween-20)**

| | |
|---|---|
| TBS | 1 liter |
| Tween-20 | 1 ml |

**Protein transfer protocol**

Twelve pieces of filter paper and a piece of nitrocellulose membrane, 0.45 μm was cut according to the size of the gel. The gel, the filter papers and the nitrocellulose membrane were immersed in the transfer buffer for 15 minutes. Six electrode papers were placed on the anode plate followed by the membrane, the gel and lastly with another six filter papers. A tube was used to roll out any trapped air bubbles. Then, the cathode tray was placed on the top of the sandwiched gel. The transfer was performed for one hour at a constant current of 0.8 $mA/cm^2$.

### 2.2.5.3 Protein detection using HRP conjugated CGB lectin

The nitrocellulose membrane was blocked using TBS-T containing 3% gelatin for one hour on a shaker at room temperature. It was then washed trice with TBS-T for 5 minutes each. The blot was incubated with HRP conjugated CGB lectin in TBS-T at an optimal concentration of 1 µg/ml at 4$^{o}$C overnight. The wash step was repeated twice. The membrane was developed using metal enhance DAB solution kit (Pierce, Illinois), prepared based on the manufacturer's protocol. The development reaction was stopped by washing the membrane with distilled water. The membrane was air dried prior to scanning.

### 2.2.5.4 Protein detection using AP conjugated CMB lectin

The procedure for the detection of CMB lectin linked glycoprotein was similar as described in Section 2.2.5.3. However, the incubation procedure was replaced with AP conjugated CMB lectin at 1:10000 dilution. To develop the membrane, the membrane was incubated in NBT/BCIP solution, which was prepared by dissolving one NBT/BCIP tablet in 10 ml of distilled water. The development reaction was stopped by washing the membrane with distilled water. The membrane was air dried prior to scanning.

## 2.2.6 Image analysis

Silver stained 2-DE gels and lectin resolved blots were scanned using an Image Scanner III (GE Healthcare Biosciences, Uppsala, Sweden). Analysis of each glycoprotein spot and protein spot was performed using Image Master Platinum 7.0 software (GE Healthcare Biosciences, Uppsala, Sweden). Percentage of volume

contribution refers to the spot volume of a protein which was expressed as a percentage of the total spot volume of all detected proteins. All values are presented as mean±S.E.M. The Student's *t*-test was used to analyze the significance of differences between normal subjects and patients. The false discovery rate control was carried out using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995).

### 2.2.7 Sample preparation for MS analysis

#### 2.2.7.1 In-gel digestion

For the protein identification by mass spectrometry, proteins resolved by 2-DE has to be digested first. Before the in-gel digestion was performed, the desired protein spot was excised manually from the modified silver-stained gels and put in a microcentrifuge tube containing deionized distilled water. The gel plugs were destained using 15 mM potassium ferricyanide in 50 mM sodium thiosulphate pentahydrate until they were transparent. They were further reduced and alkylated using 10 mM DTT in 100 mM ammonium bicarbonate and 55 mM iodoacetamide in 100 mM ammonium bicarbonate, respectively. Then, the gel plugs were washed with 50% acetonitrile in 100 mM ammonium bicarbonate and 100% acetonitrile and followed with dehydration using vacuum centrifugation. The dried plugs were incubated overnight in 25 µl of 6 ng/µl trypsin in 50 mM ammonium bicarbonate at 37$^{o}$C. Finally, the peptides were dried using a vacuum centrifugation and prepared for MS analysis.

#### 2.2.7.2 On-membrane digestion

The on-membrane digestion was done for MS analysis of peptide extracted from a nitrocellulose membrane. After electoblotting, the proteins onto a nitrocellulose

membrane, the membrane was blocked with 0.5 ml 1% PVP-40 (w/v) in TBS, washed three times with the same buffer and detected using lectins as described in Section 2.2.5. The protein spots of interest was excised carefully and destained with 0.5 ml 0.1% EDTA dihydrate for 30 minutes. Then, the nitrocellulose spots were washed three times with 0.5 ml of 0.5 and 1.0 M melibiose in 50 mM Tris-HCl pH 7.5 for 30 minutes each at room temperature to strip the horseradish peroxidase-conjugated CGB lectin. Similarly, the nitrocellulose spots were also washed three times with 0.5 ml of 0.5 and 1.0 M methyl-α-D-mannopyranoside in 50 mM Tris-HCl pH 7.5 for 30 minutes each at room temperature to strip the alkaline phosphatase-conjugated CMB lectin. Then, the stripped spot was blocked with 0.5 ml 0.5% (w/v) PVP-40 in 100 mM acetic acid at 37$^{o}$C for 30 minutes. After that, the spots were washed at least six times with deionized distilled water to remove PVP-40. The protein spots were digested with trypsin in 50 mM sodium bicarbonate at a final concentration 12.5 ng/µl and incubated overnight at 37$^{o}$C. After digestion, the samples were dried with a vacuumed centrifugation and then dissolved in acetone (22.5 µl per mm$^2$ of spot area). Then, it was vortexed and incubated for 30 minutes at room temperature. After incubation, the acetone was carefully removed and the precipitated peptides were air-dried and prepared for mass spectrometry analysis.

**2.2.7.3 Mass spectrometry analysis**

Peptides obtained from both gels and membranes were reconstituted with 0.1% formic acid and desalted using ZipTip $C_{18}$ reversed phase media (Millipore, MA, USA). The peptides were mixed with 10 mg/ml of CHCA in 1% TFA in 50:50 acetonitrile/water. Seven µl of the peptides were then spotted on the Opti-TOF 384 well insert (Applied Biosystem/MDS Sciex, Toronto, Canada). The samples were analyzed

on the 4800 Plus MALDI TOF/TOF analyzer (Applied Biosystem/MDS Scienx) with the mass standard kit (Applied Biosystem/MDS Scienx) served as the calibrator for the resulting MS and MS/MS mass spectra scales.

#### 2.2.7.4 Protein identification

Protein spots were analyzed using MASCOT search engine (Matrix Science, London, UK) against *Homo sapiens* entries in the Swiss-Prot database. The following parameters were used in the MASCOT Ion Search: 1) enzyme trypsin was used. 2) one missed cleavage was allowed. 3) fixed modification was carbamidomethyl (cystein) 4) variable modification was oxidation (methionine) 5) mass tolerance for precursor ion/peptide tolerance was 50 ppm. 6) mass tolerance for fragment ion/ms/ms tolerance was 0.1 Da. 7) Only monoisotopic mass were included in the search.

### 2.2.8 Enrichment of urinary glycoprotein using CGB and CMB lectins affinity column

**Stock solutions**

**0.1 M sodium bicarbonate pH 8.5**

| | |
|---|---|
| Sodium bicarbonate | 8.4 g |

Sodium bicarbonate was dissolved in 800 ml of distilled water and then the pH was adjusted to 8.5 and the buffer was topped up to 1 L.

**0.1 M sodium chloride**

| | |
|---|---|
| Sodium chloride | 58.44 g |

Sodium chloride was dissolved in 1 L of distilled water.

**1.0 M ethanolamine, pH 9.0**

| | |
|---|---|
| Ethanolamine | 11.8 ml |

One hundred and eighty ml of distilled water was added to 11.8 ml of ethanolamine and the pH was adjusted to pH 9.0 with HCl.

**2.2.8.1 Preparation of lectin affinity column**

Four gram of CNBr pre-activated Sepharose 4B gel was stirred overnight with 100 ml of cold distilled water at 4$^{o}$C. Twelve ml of rehydrated gel was washed with 2 liter of cold distilled water followed with 1 L 0.1 M sodium bicarbonate, pH 8.5. Then, the rehydrated gel was incubated with 12 ml of lectin solution with the concentration of 4 mg/ml under a gentle stirring for 20 hours at 4$^{o}$C. The conjugated Sepharose was then transferred into a sintered funnel and washed with 200 ml of 1.0 M sodium chloride and 200 ml distilled water. The Sepharose was suction-dried to a wet cake and then stirred in 100 ml 1.0 M ethanolamine, pH 9.0 for 1 hour at 4$^{o}$C. After the blocking step, the Sepharose was washed with 1 L of 1.0 M sodium chloride and 1 L of distilled water. Lastly, the Sepharose was poured into a column and washed with PBS buffer, pH 7.2.

**2.2.8.2 Affinity separation of glycoprotein using CGB and CMB lectin affinity column**

To prepare the sample for the affinity chromatography, 400 µg of pooled urinary protein samples from eleven control subjects, seven ECa patients, nine OCa patients and eight CCa patients were respectively dissolved with 1 ml 0.1 M PBS, pH

7.2 and loaded into four separate columns. They were left to stand overnight to allow glycoprotein binding. The first fraction prior to the washing process was collected as an unbound protein fraction. The column was washed with 0.1 M PBS, pH 7.2 and 1 ml of wash fractions were collected in a microcentrifuge tube. The entire washing process was monitored at 280 nm. The washing process was performed until the absorbance fell to the baseline. To elute the bound glycoprotein, the column was washed with the elution buffer (0.1 M melibiose for CGB lectin or 0.1 M methyl-α-D-mannopyranoside for CMB lectin in 0.1 M PBS buffer, pH 7.2). The column was left to stand overnight to ensure complete detachment of glycoproteins from the respective lectins. Fractions of eluted glycoproteins were collected in microcentrifuge tubes which were labeled as bound fractions. Regeneration of column was performed by washing the column again with 10 column bead volumes of PBS buffer followed with 5 column bead volumes of 0.1% Triton X-100 at 37$^{o}$C and lastly equilibrated with 2 ml of PBS buffer.

#### 2.2.8.3 Analysis of bound fraction with LC-MS/MS

The bound fractions were pooled, dialyzed and concentrated with Vivaspin column concentrator (Sartorius Stedim Biotech, Goettingen, Germany). The concentrated bound fractions were lyophilized and sent to The Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University, USA, for LC-MS/MS analysis.

#### 2.2.8.4 Protein identification and spectral counting

The LC-MS/MS data were viewed and analyzed using Scaffold version 3.4.5 software. The protein identification was done by submitting the raw MS/MS data to the MASCOT search engine (Matrix Science, London, UK) against *Homo sapiens* entries in

the Swiss-Prot database. The following parameters were used to match the peptide: 1) enzyme trypsin was used. 2) one missed cleavage was allowed. 3) fixed modification was propionamide 4) variable modification was oxidation (methionine) and acetylation 5) mass tolerance for precursor ion/peptide tolerance was 20 ppm. 6) mass tolerance for fragment was 1.0 Da. The identity of a peptide was considered valid when XCorr value was greater than 0.8 for singly charged ions, 1.9 for doubly charge ions and 2.6 for triply charged ions and have a minimum peptide length of seven amino acids. The subcellular location and the glycosylation type of the identified proteins were searched from the Swiss-Prot database. Additionally, the potential *O*-or *N*-glycosylation of the identified protein were determined with NetOGlycan 3.1 (http://www.cbs.dtu.dk/services/NetOGlycan) and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). The spectral of the proteins were normalized with Normalized Spectral Abundance Factors (NSAF) and subjected to the comparison between control and cancer groups (Florens *et al.*, 2006; Zybailov *et al.*, 2006). NSAFs were used to estimate the protein level, in which the values closer to 1 indicating higher protein level.

$$(\mathrm{NSAF})_k = \frac{(\mathrm{SpC}/\mathrm{Length})_k}{\sum_{i=1}^{N} (\mathrm{SpC}/\mathrm{Length})_i}$$

The NSAF for a protein k is the number of spectral counts of a protein (k) divided by the protein's length, divided by the total number spectral counts for all N proteins. The ratio of change was calculated by dividing the NSAF value of cancer cohort over the NSAF value of control. The ratio value greater or equal to 1.6 was considered increase in expression of a protein, while the ratio value lesser or equal to 0.67 was considered decrease in expression of a protein.

### 2.2.9 Competitive ELISA

**Standard solutions**

**Binding buffer**

**(0.05 M sodium bicarbonate buffer, pH 9.6)**

| | |
|---|---|
| Sodium carbonate | 1.59 g |
| Sodium hydrogen carbonate | 2.93 g |

Both salts were dissolved in 1 L of distilled water

**Blocking buffer: 0.5% gelatin in PBS-T**

**Washing buffer: PBS with 0.05% (v/v) Tween-20**

**Stopping solution: 2M sulphuric acid ($H_2SO_4$)**

**ELISA protocols**

Ten μg of freeze-dried urine samples were dissolved with 50 μl of binding buffer and loaded to microtitre plates (Nunc, Denmark) for coating overnight at 4 $^{o}$C. The wells were rinsed extensively with 100 μl of washing buffer for three times. Non-coated sites were blocked with 100 μl of blocking buffer for 30 minutes in room temperature. After blocking, the plates were washed and incubated with mouse anti-human CLU (BioRad Laboratories Inc., U.S.A.) (MCA2612, 1.0 mg/ml) in the presence of controls' and patients' urines at dilution 1:1000 in PBS-T for 1 hour at room temperature. Positive controls were prepared in the absence of urine samples. Blanks were prepared by addition of washing buffer.

After incubation, each wells were washed extensively with washing buffer and followed by additional of 200 µl of HRP labelled secondary anti-mouse IgG at dilution 1:5000. Plates were incubated for another 1 hour at room temperature and then washed with washing buffer. One hundred µl of 3,3',5,5'-tetramethylbenzidine substrate (Sigma Aldrich Company, St. Louis, U.S.A) were added to each wells and incubated for 15 minutes at room temperature in the dark. The reaction was terminated by the addition of 100 µl stopping buffer after the incubation period. Absorbance values were read at 450 nm using an ELISA plate reader. The amount of CLU in urine samples was determined with the percentage of inhibition of substrate hydrolysis. The formula is percentage of inhibition = [1-(sample absorbance/positive control absorbance)] X 100 (Reddington *et al.,* 1991).

### 2.2.10 Lectin-ProteinChip array for SELDI-TOF analysis

**Standard solutions**

**Trypsin denaturation solution**

**(8 M urea, 50 mM Tris-HCl, 2 mM DTT, pH 8)**

| | |
|---|---|
| Urea | 4.8 g |
| Tris base | 61 mg |
| DTT | 3.09 mg |

The chemicals were dissolved in 5.0 ml and the pH of the solution was adjusted using 1.0 M HCl. The solution was topped up to 10 ml with distilled water.

**Coupling buffer: PBS, pH 7.5**

**Blocking buffer: 1 M Ethanolamine in PBS, pH 7.5**

| | |
|---|---|
| 10 X PBS | 10 ml |
| Deionised distilled water | 70 ml |
| Ethanolamine | 11.8 ml |

The solution was mixed and the pH of the solution was adjusted to pH 7.5 with 1.0 M NaOH. The solution was topped up to 100 ml with distilled water.

**Wash buffer: 0.5% Triton-X in PBS**

**Matrix: α-cyano-4-hydroxycinamic acid (CHCA)**

**2.2.10.1 Sample preparation**

Freeze-dried urine sample, which was stored in -80$^{o}$C freezer prior to SELDI-TOF analysis, was reconstituted with 100 µl of 50 mM ammonium bicarbonate. Then, 5 µl of 200 mM DTT stock solution was added to the solution. The solution was boiled for 10 minutes at 95$^{o}$C before 4 µl of 1 M iodoacetamide was added to the solution. After the mixture was vortexed and followed by a quick spin, it was incubated for 1 hour at room temperature. Then, the solution was added with 20 µl of DTT, vortexed, spun and incubated for another 1 hour at room temperature. The sample was incubated with 10 µl of 100 ng/µl trypsin in 50 mM ammonium bicarbonate overnight at 37$^{o}$C. After digestion, 25 µl of trypsin denaturation solution was added to the solution and the mixture was boiled at 95$^{o}$C for 15 minutes. Finally, the solution was dried with vacuum

centrifugation. The peptides were reconstituted with 0.1% formic acid and carefully desalted using ZipTip $C_{18}$ reversed phase media (Millipore, MA, USA). Lastly, it was let to dry at room temperature.

### 2.2.10.2 Immobilization of CGB and CMB lectin to pre-activated array

The immobilizations of CGB and CMB lectins were performed as described by the manufacturer with some modifications. Firstly, each spot was wetted with 2 µl of 50% acetonitrile/deionized distilled water briefly and then the solution was removed with the edge of a kimwipe. The spot was then activated by adding 2 µl of PBS and the chip was incubated in a humidity box for 15 minutes at room temperature. The PBS was then removed with the edge of a kimwipe. Next, lectin immobilization was performed by adding 5 µl of 1.0 mg/ml of lectin in 0.1 M sodium bicarbonate buffer (pH 8.0) on the spot and the chip was incubated for 1 hour at room temperature. After the incubation period, the lectin solutions were removed from the spots. The residual active sites on the spots were blocked with 5 µl of blocking buffer and it was incubated for 30 minutes. The spots were then washed twice with washing buffer and three times with PBS by pipetting the buffer up and down for 15 times. The samples were prepared as described as Section 2.2.10.1. It was mixed with 5 µl of washing buffer and added onto spots and incubated overnight at $4^{o}$C. After incubation, the spots were washed twice with 5 µl of wash buffer, twice with PBS and finally rinsed with MiliQ water. One µl of matrix was added to each spot and mass analysis was performed in a PCS4000 ProteinChip reader, Enterprise edition (BioRad Laboratories Inc., U.S.A.) with a combination of 130 laser shots at optimum laser energy at 1500 nJ for each spot. Each sample was analyzed in duplicate.

**2.2.10.3 Data analysis of lectin captured glycopeptides**

The raw data were analyzed with ProteinChip Data Manager Software 4, Enterprise edition (BioRad Laboratories Inc., U.S.A.). Peak detection was performed by Expression Difference Mapping which involved the baseline subtraction and automatic peak detection. The peaks were detected between 1 kDa to 10 kDa and generated a list of peak clusters using the setting of five times to the signal-to-noise (S/N) ratio for the first pass and and two times to the S/N ratio for second pass. The statistical analysis using Wilcoxson signed-rank test was subsequently performed for all peak intensity values of control and three cohorts of cancer. A *p*-value of less than 0.05 was considered to be statistically significant. To enable the peak clusters to be analyzed by Biomarker Pattern Software 5.0, the expression difference mapping (EDM) file was initially converted to csv. file. The biomarker pattern of control and three cohorts of cancers was identified which subsequently was used to construct the diagnostic model of CGB or CMB lectin captured urinary glycopeptides profile.

# Chapter 3
# Results

# CHAPTER 3: RESULTS

## 3.1 Isolation and purification of Champedak Galactose-Binding (CGB) and Champedak Mannose-Binding (CMB) lectins

The crude extract of champedak seeds was prepared as described in Section 2.1.1. The CGB and CMB lectins were then isolated from the crude seed extract by affinity column chromatography using immobilized galactose and mannose to capture the CGB and CMB lectins, respectively (Section 2.2.1). Figures 3.1a and 3.1b demonstrate the typical profiles generated from elution of the CGB and CMB lectins by the respective sugar affinity columns. Both lectins appear to have been eluted as single sharp peaks, which is a reflection of the purity of the eluted proteins.

## 3.2 Determination of the concentrations of CGB and CMB lectins

Estimation of protein concentrations of 5 ml CGB and CMB lectin extracts was performed using the BCA protein assay kit (Pierce, Rockford, Illinois). A standard calibration curve of absorbance values over several concentrations of BSA was constructed. Figure 3.2 shows a typical calibration curve obtained from the protein assay. Based on the extrapolated standard curve, the concentrations of the CGB and CMB lectins were approximately 720 µg/ml and 260 µg/ml, respectively. In addition, the yield per gram of CGB and CMB lectins' wet weight were 3.5 mg and 1.2 mg, respectively.

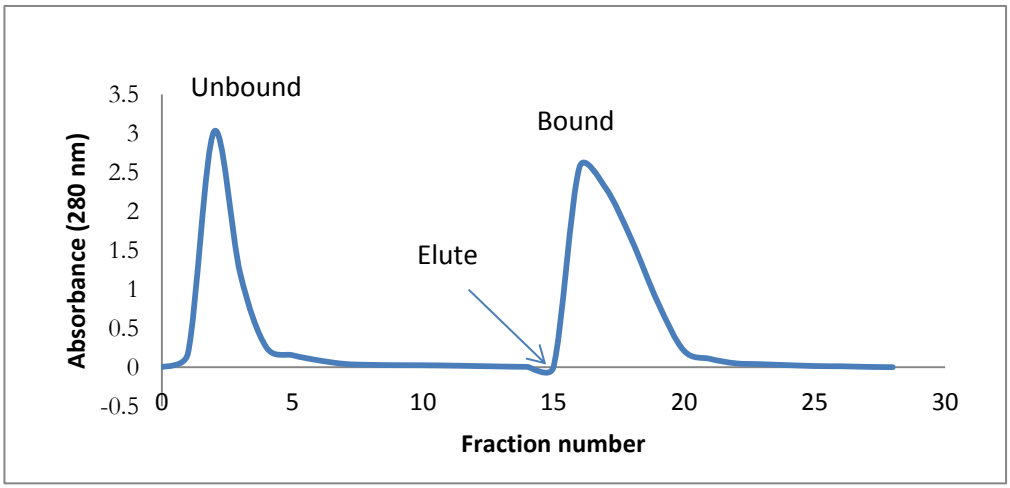
## 3.3 Determination of purity of CGB and CMB lectins

The affinity isolated lectins were subjected to SDS-PAGE electrophoresis using 18% separating gel to confirm for their purity. Figure 3.3 shows the SDS-PAGE

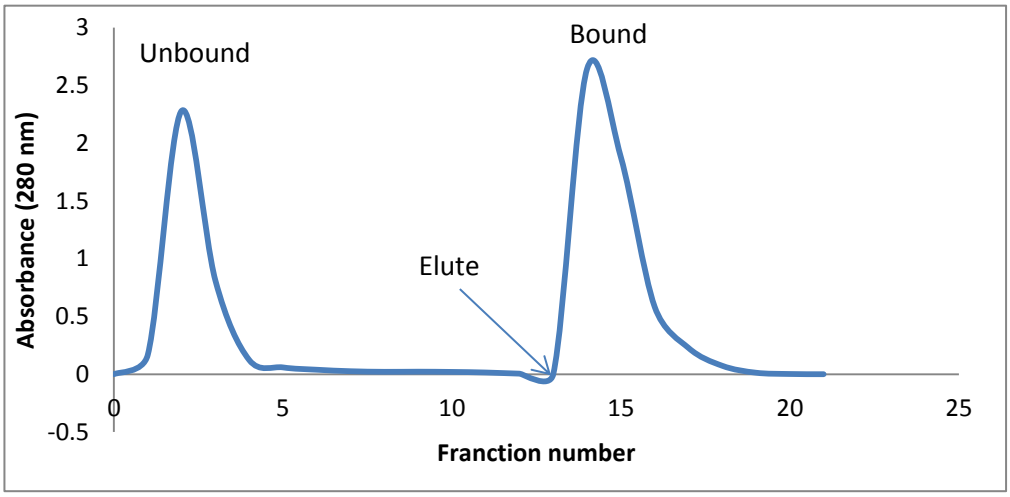
**Figure 3.1: Typical elution profiles of CGB and CMB lectin**

Panel A demonstrates the elution profile of CGB lectin when crude lectin was subjected to the immobilized galactose-Sepharose affinity column chromatography. Fractions collected from the galactose column were monitored at absorbance of 280 nm. The unbound fractions with high absorbance were kept for CMB lectin isolation. Washing was performed using PBS, pH 7.2 to eliminate unbound substances until a baseline reading was achieved. CGB lectin was eluted with 0.8 M D-galactose in PBS buffer and arrow indicates the start of elution. Panel B demonstrates the isolation of CMB lectin from the unbound fractions of galactose-Sepharose column using mannose-Sepharose column. Arrow indicates the start of elution for CMB-lectin using 0.8 M D-mannose in PBS buffer.

A


Unbound
Bound
Elute
Absorbance (280 nm)
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
0
5
10
15
20
25
30
Fraction number

B


Unbound
Bound
Elute
Absorbance (280 nm)
3
2.5
2
1.5
1
0.5
0
-0.5
0
5
10
15
20
25
Franction number

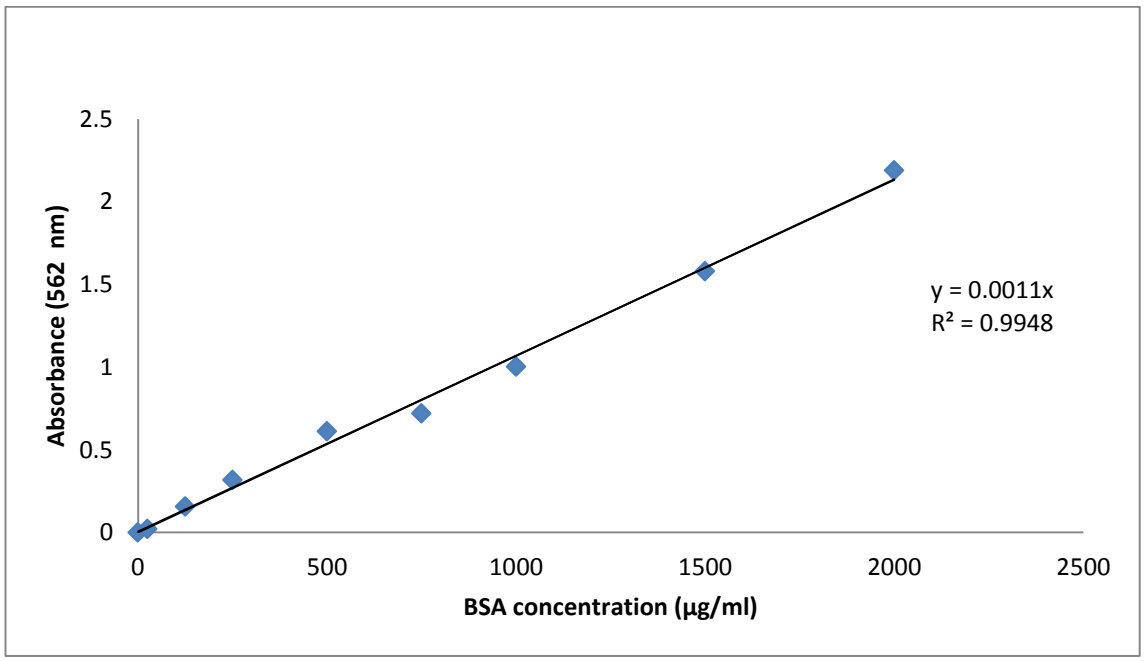



**Figure 3.2: Standard curve for protein concentration estimation**

A typical standard curve was constructed based on serial dilutions of BSA. BSA was diluted to obtain protein concentrations ranging from 0 to 2 mg/ml. Estimation of CGB and CMB lectin concentrations was made by referring to a standard curve. Absorbance of the complex formed was measured at 562 nm.

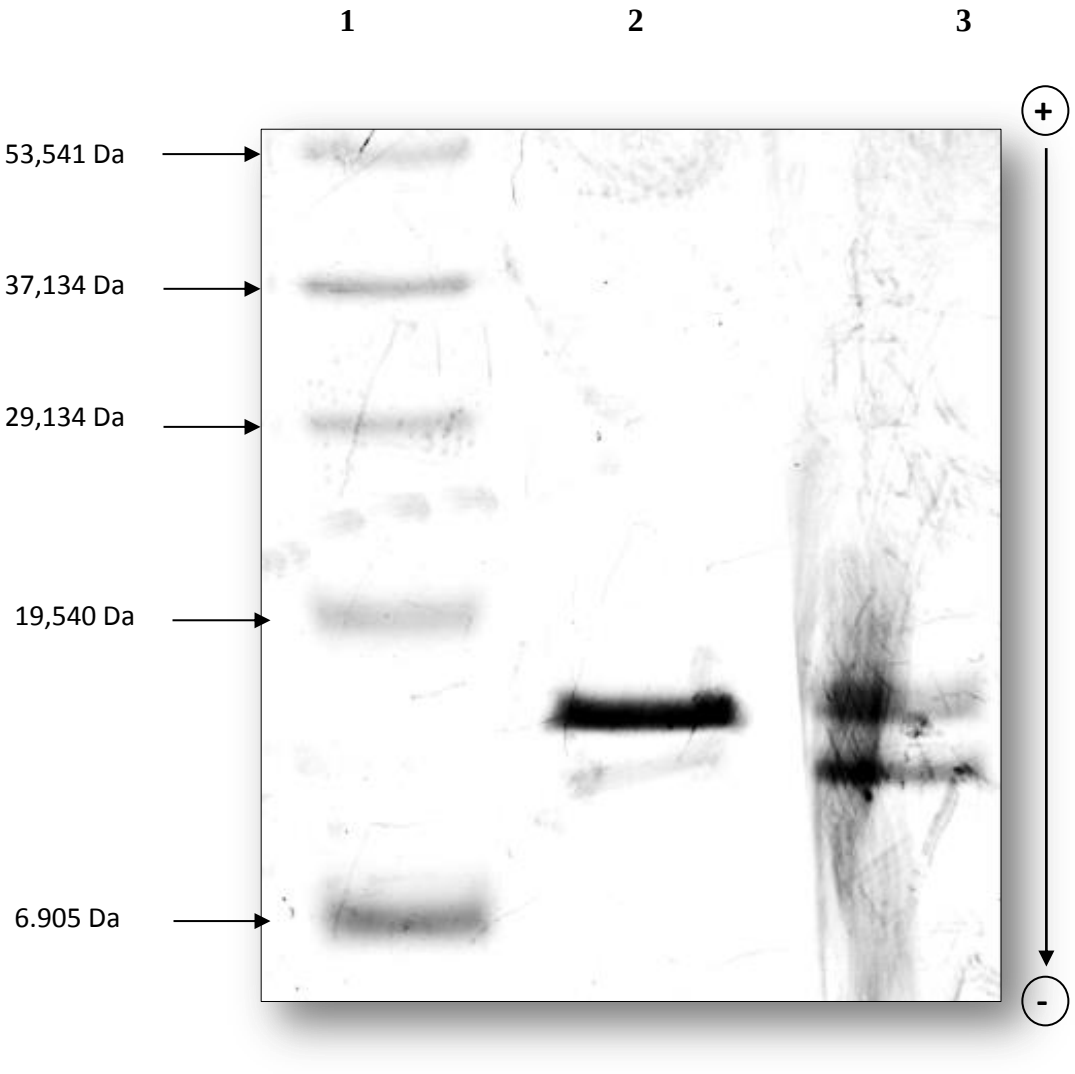



**Figure 3.3: SDS-PAGE analysis of purified CGB and CMB lectin**

The purified CGB and CMB lectins were subjected to 18% SDS-PAGE, followed by Coomassie Brilliant blue staining to check its purity. Under reducing condition, CGB lectin was resolved into two bands with M*r* of approximately 14,500 Da and 16,000 Da while CMB was resolved into only one band with M*r* of approximately 15,000 Da.

Lane 1: Pre-stained SDS-PAGE marker consisting of ovalbumin (53,541 Da), carbonic anhydrase (37,134 Da), soybean trypsin inhibitor (29,134 Da). lysozyme (19,540 Da) and aprotinin (6,905 Da)

Lane 2: CMB lectin

Lane 3: CGB lectin

analysis of the affinity purified CGB- and CMB-lectins. Electrophoresis was performed under reducing conditions as described in Section 2.2.3.4. Under this condition, the CGB lectin was resolved into two bands with relative molecular weights of approximately 14,500 and 16,000 Da. The CMB lectin, on the other hand, was resolved into a single band with relative molecular weight of approximately 15,000 Da. These bands were the characteristic bands of the CGB and CMB lectins that were earlier established (Lim *et al.,* 1997; Abdul Rahman *et al.,* 2002). There were no other bands observed in this SDS-PAGE gel, indicating that the lectins were pure.

## 3.4 Determination of urinary protein concentration

The concentrations of urinary protein samples were determined using a similar method as that used for the determination of the concentrations of the lectins. The control urine samples apparently contain lower amounts of proteins compared to the urine samples of the groups of patients with three different types of cancers (Table 3.1).

**Table 3.1: Concentration of protein in urine samples**

| Urine samples | Concentration of urinary protein (µg/ml) |
|---|---|
| Control (n=11) | 101.302 ± 10.670 |
| ECa (n=7) | 355.292 ± 106.420* |
| OCa (n=9) | 648.215 ± 126.840* |
| CCa (n=8) | 372.837 ± 90.230* |

All values were expressed as mean±S.E.M.
*statistically significant compared to control when $p<0.05$.

## 3.5 Two-dimensional electrophoretic analysis

### 3.5.1 2-DE urinary protein profiles of controls and patients with ECa

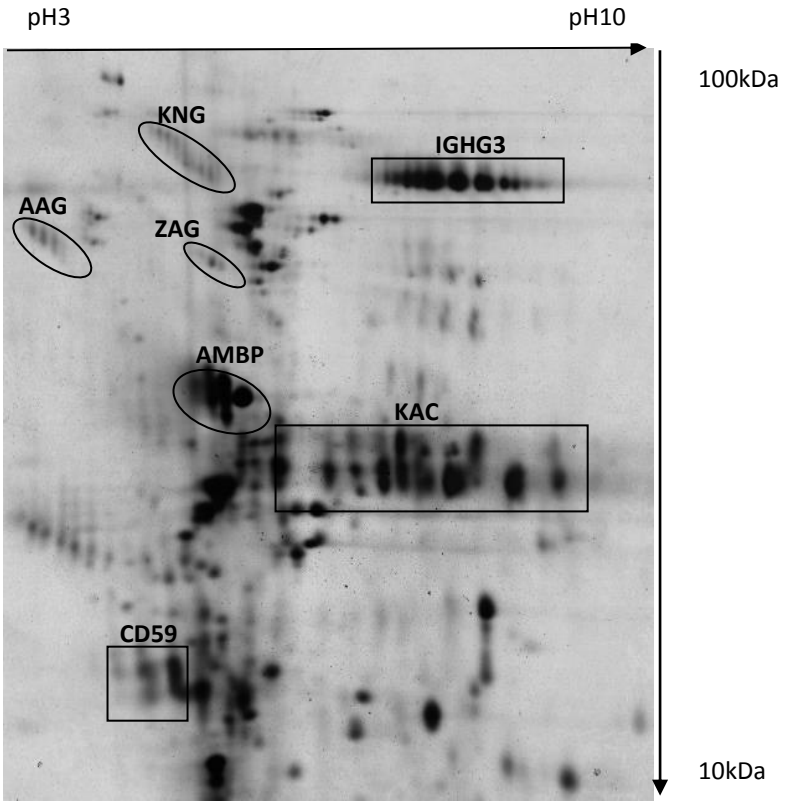
The urinary proteins of the control female subjects were subjected to 2-DE followed by silver staining. Comparable 2-DE urinary protein profiles were generally obtained for all samples analysed. Figure 3.4a shows a representative 2-DE urine profiles of controls and patients with ECa. There were seven protein clusters which were highly resolved and consistently appeared in all the proteome maps of the controls. Furthermore, the intensities of these protein clusters were not significantly different between the 11 control samples analysed. At this stage, preliminary identification of protein spots was performed by visually comparing the 2-DE profiles of the urinary proteins that were generated in this study with the earlier established 2-DE urine proteome map of Candiano et al. (2010).

Since the 2-DE urinary protein profiles of patients with OCa and CCa have been earlier established by Abdullah-Soheimi et al. (2010), this part of the study focused solely on patients with ECa. When the urinary proteins from seven patients with ECa who were newly diagnosed and untreated were resolved with 2-DE, a similar pattern of protein distribution was obtained for all the seven proteins that consistently appeared in the 2-DE gels. Proteins that were resolved in both control and ECa groups samples were kininogen-1 (KNG), alpha-1 acid glycoprotein (AAG), zinc alpha-2 glycoprotein (ZAG), CD59, protein AMBP (AMBP), Ig gamma 3 chain C region (IGHG3) and Ig kappa chain C region (KAC).

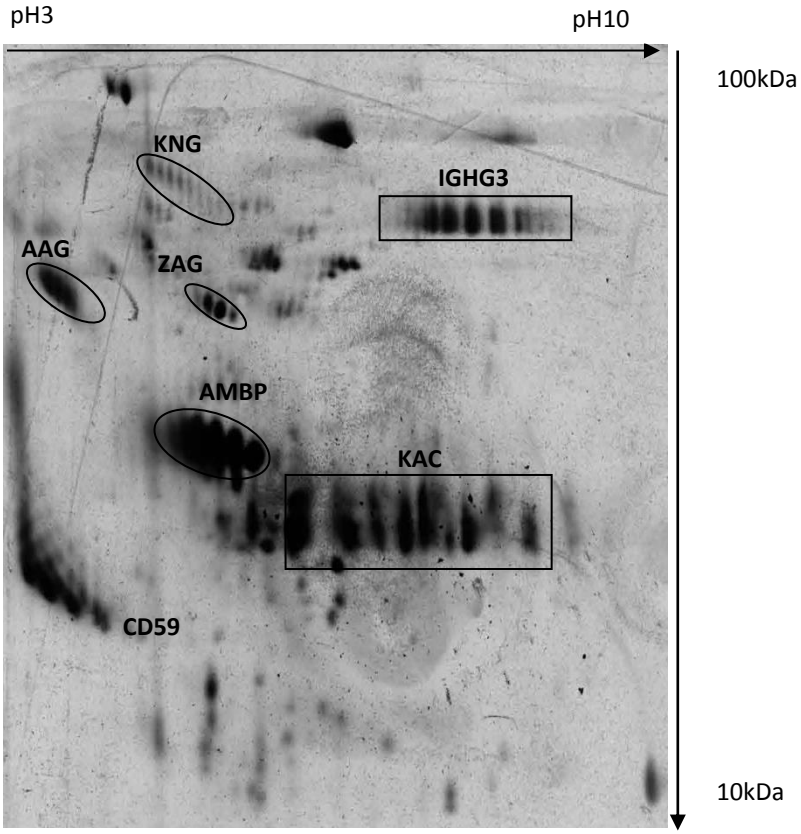
**Figure 3.4: Typical 2-DE urine profiles of controls and patients with ECa**

Urinary proteins from controls and patients with ECa were separated according to their isoelectric points and molecular mass. Panel A and panel B demonstrate representative 2-DE urinary protein profiles of controls and patients with ECa, respectively. The labeled protein clusters are proteins which were consistently observed in both groups. The acidic side of the gel is to the left and relative molecular mass declines from the top.

**A: Control**




**B: ECa**



### 3.5.2 Image analysis of 2-DE urinary protein profile of controls and ECa patients

The silver-stained 2-DE urinary protein profiles of controls and patients with ECa were scanned and analyzed using the Image Master Platinum Software version 7.0. All well-resolved protein spots were estimated for their percentage of volume contribution. Volume analysis of urinary proteins in the 2-DE profiles of controls and patients demonstrated the similar results for some of the proteins, except for AAG, ZAG and CD59. The intensities of AAG and ZAG were more than 10-fold higher while the intensity of CD59 protein clusters was much lower in the 2-DE urinary profiles of patients with ECa. These values were found to be significantly different between both groups when analyzed with the Student *t*-test followed by the false discovery rate control, based on the method of Benjamini and Hochberg. Table 3.2 shows the mean percentage of volume contribution of the seven urinary proteins that consistently appeared in controls and patients with ECa.

**Table 3.2 Mean percentage of volume contribution of urinary proteins**

| Urinary proteins | Mean % Vol ± SEM | | *P* | Fold changes* |
|---|---|---|---|---|
| | Control | ECa | | |
| AAG | 0.161 ± 0.072 | 2.746 ± 0.717 | 0.001 | +13.03 |
| ZAG | 0.175 ± 0.045 | 2.184 ± 0.592 | 0.001 | +10.59 |
| CD59 | 2.575 ± 0.497 | 0.177 ± 0.070 | 0.002 | –14.55 |
| KNG | 5.137 ± 1.826 | 0.945 ± 0.491 | 0.108 | ns |
| IGKC | 14.785 ± 2.197 | 18.840 ± 2.651 | 0.286 | ns |
| AMBP | 8.941 ± 1.706 | 12.837 ± 3.430 | 0.306 | ns |
| IGHG3 | 4.014 ± 1.221 | 4.554 ± 1.706 | 0.807 | ns |

*Fold expression changes are relative to the control values; (+) increase in expression; (–) decrease in expression; ns – not statistically significant; A *p*-value of less than 0.0214 was considered significant.

## 3.6 Profiling of urinary glycoproteins of controls and patients with ECa, OCa and CCa using 2-DE and lectins

### 3.6.1 CGB lectin detected 2-DE urinary glycoprotein profile

To obtain profiles comprising only *O*-glycosylated urinary glycoproteins, the 2-DE separated urinary proteins of patients and controls were transferred onto the nitrocellulose membrane and probed with enzyme-conjugated CGB lectin. The glycoprotein profile obtained was then compared to the earlier established 2-DE urine proteome map of Candiano et al. (2010) as well as the silver-stained 2-DE urinary protein profile that was earlier described in this study (Section 3.5.1). The *O*-glycosylated urinary protein profiles developed with enzyme-conjugated CGB lectin were entirely different from the 2-DE urinary protein profiles generated by silver staining. Six clusters of *O*-glycoproteins were consistently detected in the urinary membrane profiles of controls and patients' samples. In addition, one *O*-glycoprotein spot, with a molecular mass of approximately 51 kDa, seemed to appear in the profiles of all the control samples that were analyzed but faintly appeared in only one of the urine samples from patients with ECa. Six clusters of *O*-glycosylated urinary glycoproteins were consistently detected in all *O*-glycosylated urinary glycoproteome profiles which were Ig alpha-1 chain C region (IGHA1), KNG, inter alpha trypsin inhibitor heavy chain H4 (ITIH4), AMBP, heparan sulfate proteoglycan 2 (HSPG2) and transgelin (TAGLN). In addition, two *O*-linked urinary glycoprotein spots or clusters of hemopexin (HPX) and leucine rich alpha-2 glycoprotein (LRG) appeared with low intensity in the *O*-glycosylated urinary glycoproteome profiles of some controls or patients samples. There were two additional protein spots or clusters that were not detected in the 2-DE silver stained profiles but were clearly resolved in the *O*-glycosylated urinary glycoproteome profiles. They were identified as nebulin (NEB)

and CLU with mass spectrometry analysis. NEB appeared in the profiles of all the samples of controls and samples of patients with OCa and CCa but faintly appeared in samples of patients with ECa. At the same time, CLU consistently appeared in the profiles of all samples of the patients with OCa. Figure 3.5 shows the typical *O*-glycosylated urinary glycoproteome profiles for controls and patients with ECa, OCa and CCa.

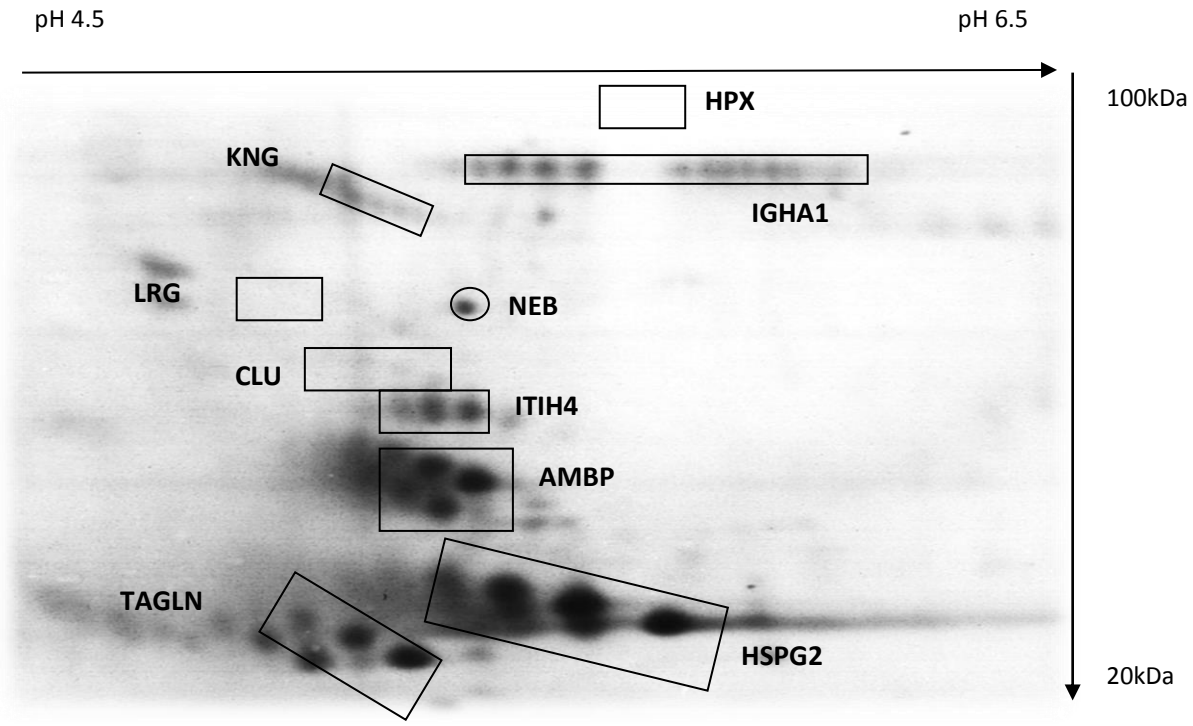
### 3.6.2 Image analysis of CGB lectin interacted 2-DE urinary glycoprotein profile

The *O*-glycosylated urinary glycoproteome profiles of controls and patients with ECa were subjected to image analysis with Image Master 7.0 Platinum and statistical analysis. The analysis showed a 51 kDa protein spot, known as NEB, was significantly different with 10.6-fold higher in the controls compared to patients with ECa. Similar analysis was performed on the *O*-glycosylated urinary glycoproteome profiles of patients with OCa. Higher levels of LRG and CLU were observed in all samples of patients with OCa, although the proteins were earlier reported to not to be differentially expressed from a previous proteomic study (Abdul-Soheimi *et al.*, 2010). Nevertheless, low levels of KNG detected in the urine of OCa patients was confirmed. Table 3.3 shows the relative level of NEB, LRG, CLU and KNG. On the other hand, no significant changes were observed in *O*-glycosylated urinary glycoproteome profiles from patients with CCa when their level was compared to that of the controls. Figure 3.6 shows the mean percentage of volume contribution of CGB lectin detected urinary glycoproteins from controls and patients with ECa, OCa and CCa.

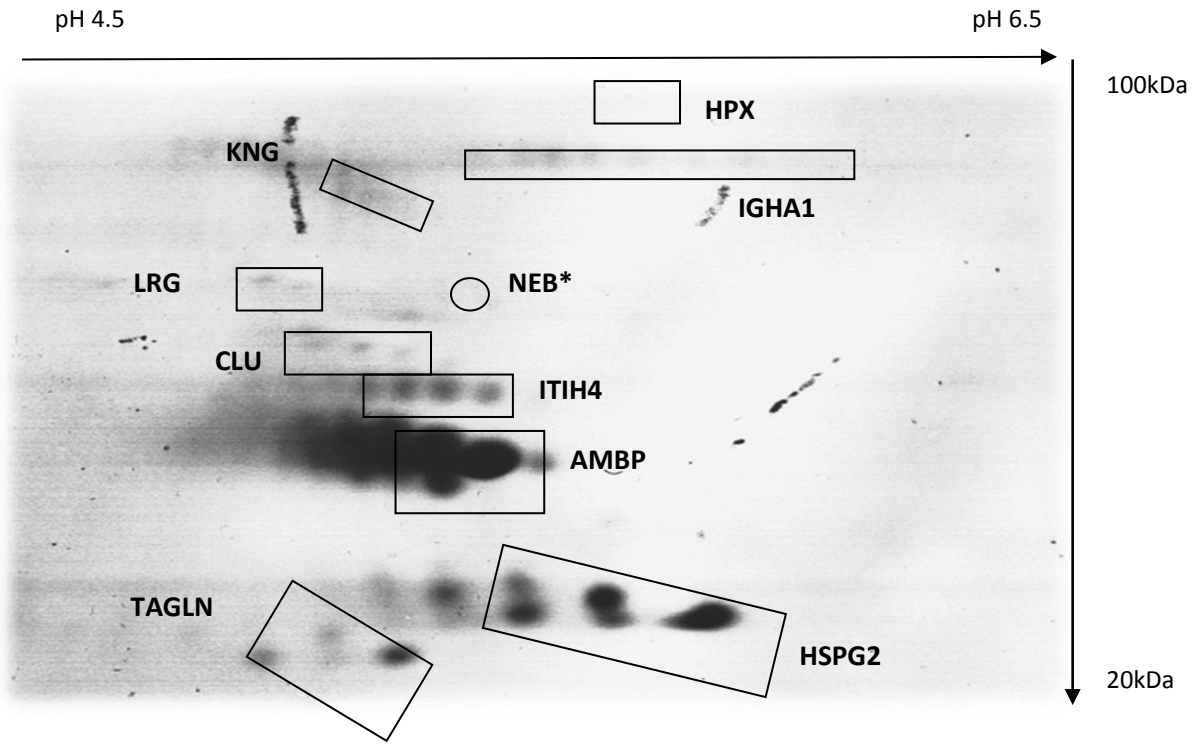
**Figure 3.5: Typical *O*-glycosylated urinary glycoproteome profiles**

Panels A, B, C and D refer to typical *O*-glycosylated urinary glycoproteome profiles of the control, patients with ECa, OCa and CCa, respectively. The labeled protein clusters refer to proteins which were consistently detected by CGB lectin using Western Blotting approach. * indicates proteins with intensity which were significantly different relative to that of control. Acidic side of the membrane is to the left and relative molecular mass declines from the top.

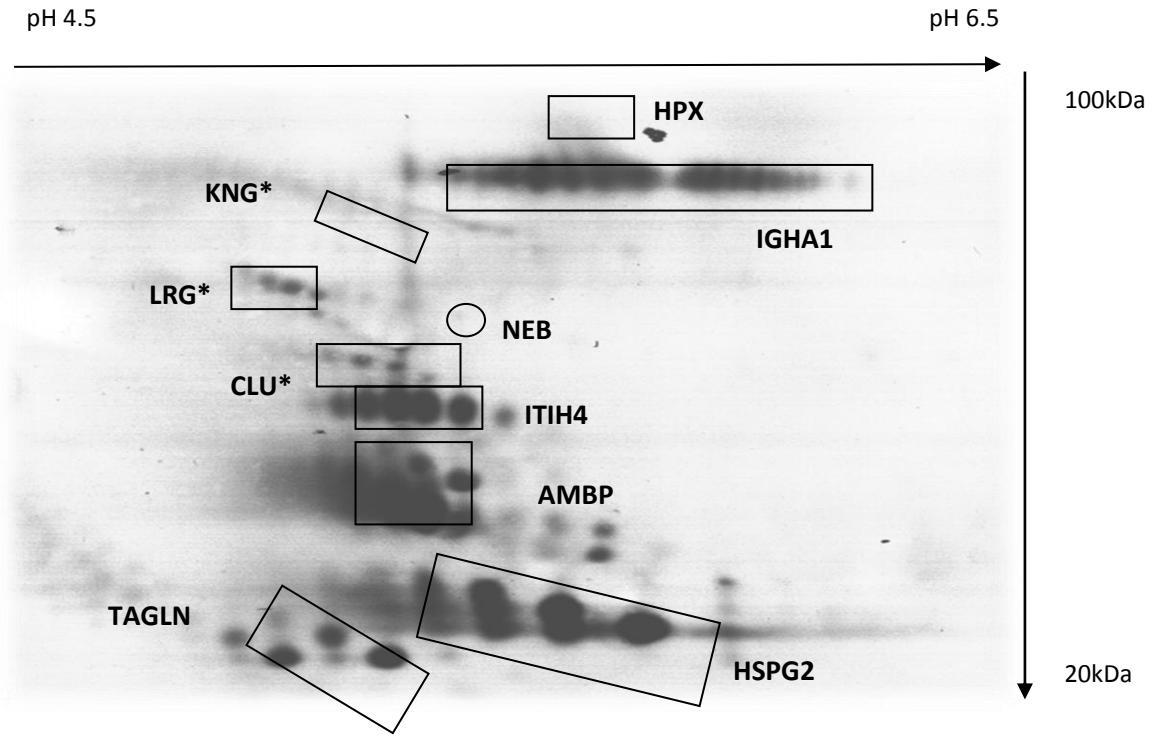
**A: Control**




**B: ECa**



**C: OCa**


pH 4.5
pH 6.5
100kDa
20kDa
HPX
KNG*
IGHA1
LRG*
NEB
CLU*
ITIH4
AMBP
TAGLN
HSPG2

**D: CCa**

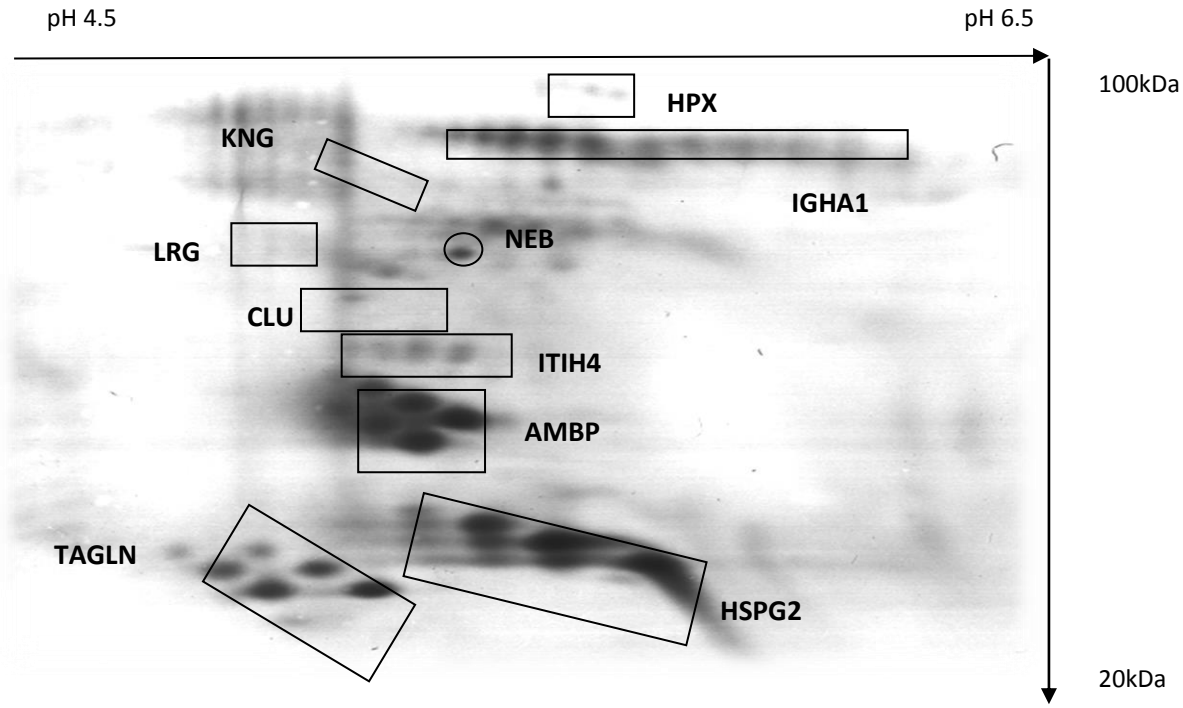

pH 4.5
pH 6.5
100kDa
20kDa
HPX
KNG
IGHA1
LRG
NEB
CLU
ITIH4
AMBP
TAGLN
HSPG2

**Figure 3.6: Mean percentage of volume contribution of CGB lectin detected urinary glycoproteins from controls and patients with ECa, OCa and CCa**

The percentages of volume contribution of each urinary glycoproteins for control (n=11), ECa (n=7), OCa (n=9) and CCa (n=8) were calculated using Image Master 2D Platinum Software 7.0. The values are expressed as mean of the percentages of volume contribution ± S.E.M. Student *t*-test was used to analyze the significance of difference between controls and the patients. False discovery rate was performed using the method of Benjamini and Hochberg. Proteins marked with * were significantly different compared to that of control. X-axis represents the sample group and Y-axis represents the percentage of volume contribution.

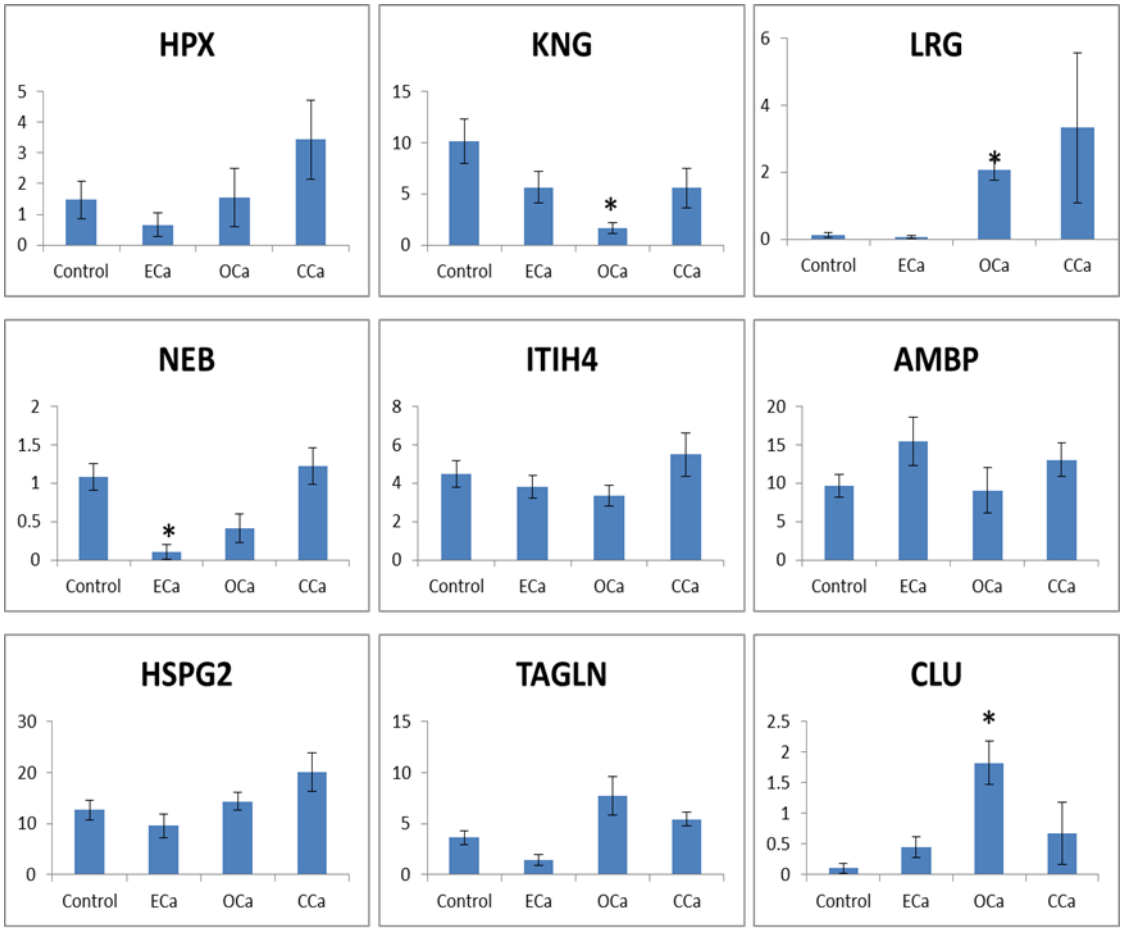

HPX
5
4
3
2
1
0
Control
ECa
OCa
CCa
KNG
15
10
5
0
*
Control
ECa
OCa
CCa
LRG
6
4
2
0
*
Control
ECa
OCa
CCa
NEB
2
1.5
1
0.5
0
*
Control
ECa
OCa
CCa
ITIH4
8
6
4
2
0
Control
ECa
OCa
CCa
AMBP
20
15
10
5
0
Control
ECa
OCa
CCa
HSPG2
30
20
10
0
Control
ECa
OCa
CCa
TAGLN
15
10
5
0
Control
ECa
OCa
CCa
CLU
2.5
2
1.5
1
0.5
0
*
Control
ECa
OCa
CCa

**Table 3.3: The relative level of *O*-glycosylated urinary glycoprotein**

| ***O*-glycosylated urinary glycoprotein** | **Fold changes** | | |
|---|---|---|---|
| | **ECa** | **OCa** | **CCa** |
| KNG | n.s | -6.19 | n.s |
| NEB | -10.59 | n.s | n.s |
| LRG | n.s | +17.17 | n.s |
| CLU | n.s | +17.34 | n.s |

Fold changes were relative to that of control
(-) decrease in level; (+) increase in level
n.s: not significant

### 3.6.3 CMB lectin detected 2-DE urinary glycoprotein profile

When 2-DE separated urinary proteins from samples of controls were probed with enzyme-conjugated CMB lectin, different profiles consisting of exclusively *N*-glycosylated urinary proteins were obtained. The CMB lectin interacted urinary glycoproteins profiles were entirely different from those generated using the CGB lectin. More spots were apparently detected by the CMB lectin detection method. The *N*-glycosylated urinary proteins were initially identified by comparison with the earlier established 2-DE silver-stained urinary profiles (Section 3.5.1) as well as published 2-DE urinary proteome (Candiano *et al.*, 2010). The *N*-glycosylated urinary proteins that consistently appeared in the 2-DE profiles of the controls were HPX, AAG, ZAG, IGHA1, KNG, ITIH4, AMBG, HSPG2,TAGLN, apolipoprotein D (APOD) and retinol binding protein 4 (RBP4).

When the analysis was performed on the 2-DE separated urinary proteins from patients with ECa, OCa and CCa who were newly diagnosed and untreated, similar *N*-glycosylated urinary protein profiles were obtained, with the exception of several

protein spots. LRG and haptoglobin β chain (HPβ) were consistently detected in all CMB blots of the OCa patients and a considerable number of CMB blots of patients with CCa (n=4) and ECa (n=5). Like the previous CGB blots, the CMB blot also unmasked CLU, which could not be detected in the 2-DE silver-stained urinary profiles. Figure 3.7 shows the typical *N*-glycosylated urinary protein profiles for controls and patients with ECa, OCa and CCa.

### 3.6.4 Image analysis of CMB lectin interacted 2-DE urinary glycoprotein profile

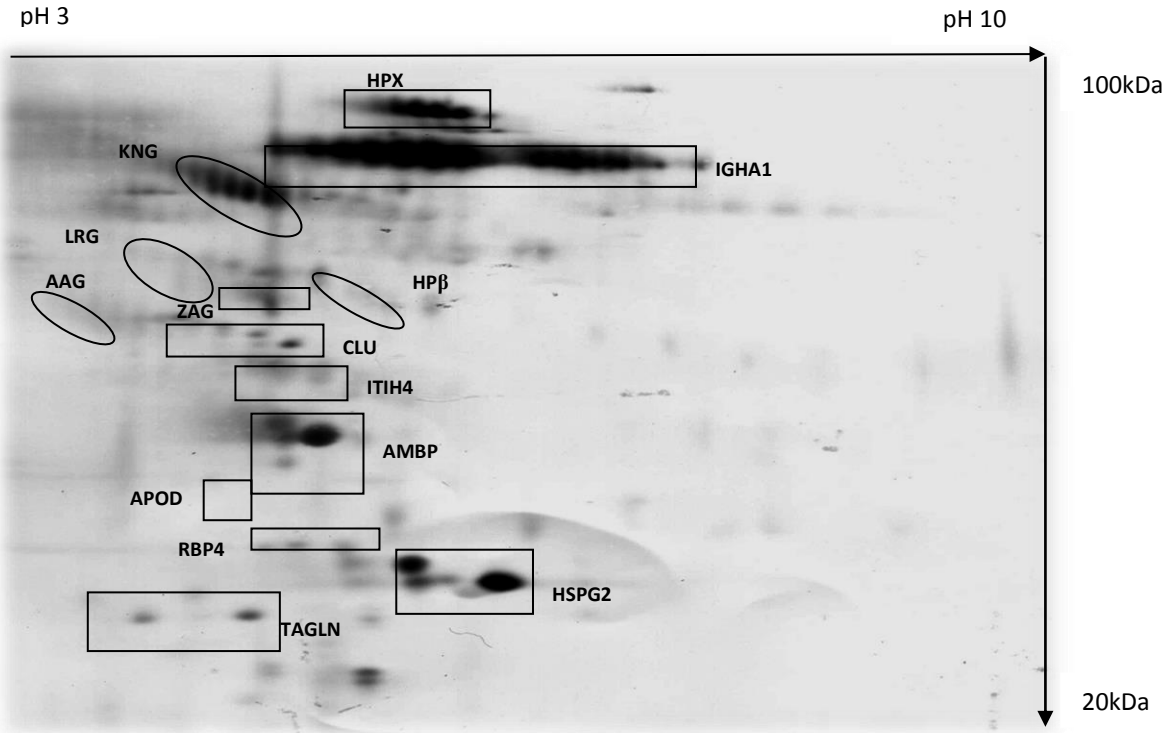
When the *N*-glycosylated urinary glycoproteins profiles of controls and patients with ECa were scanned and subjected to image analysis using Image Master Platinum 7.0 software, higher levels of AAG and ZAG were detected in the patients. This data provides confirmation of the differential levels of the proteins that was earlier established in patients with ECa using the silver-staining approach (Section 3.5.1).

A similar comparative analysis conducted on the CMB lectin blots from patients with OCa and controls also revealed the overexpression of urinary AAG and ZAG in the patients. In addition to these two protein clusters, LRG, HPβ and CLU were also demonstrated to be enhanced in the CMB lectin blots from patients with OCa. However, an earlier work by silver staining performed on patients with OCa had reported that these urinary glycoproteins were found not significantly different relative to the controls (Abdullah-Soheimi *et al.*, 2010). On the other hand, the significantly low level of urinary KNG detected in the patients with OCa was validated in this study.

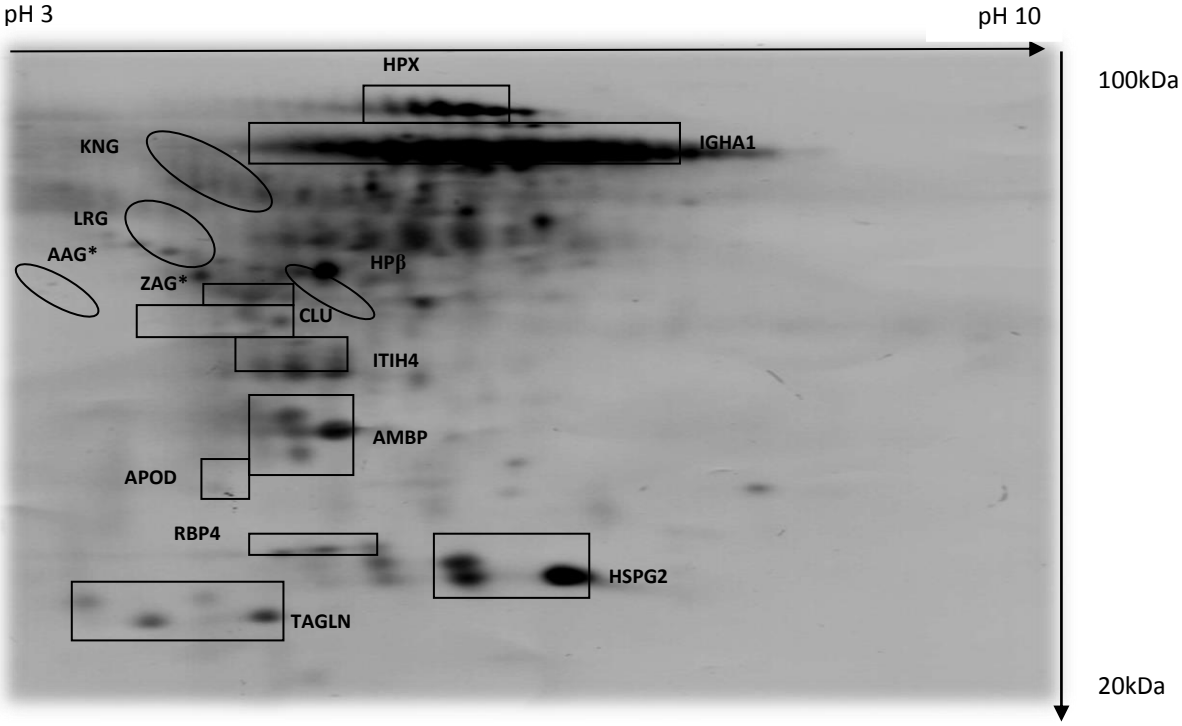
**Figure 3.7: Typical *N*-glycosylated urinary glycoproteome profiles**

Panels A, B, C and D refer to typical *N*-glycosylated urinary glycoproteome profiles of the control, patients with ECa, OCa and CCa, respectively The labeled protein clusters refer to proteins which were consistently detected by CMB lectin. * indicates proteins with intensity which were significantly different relative to that of control. Acidic side of the membrane is to the left and relative molecular mass declines from the top.

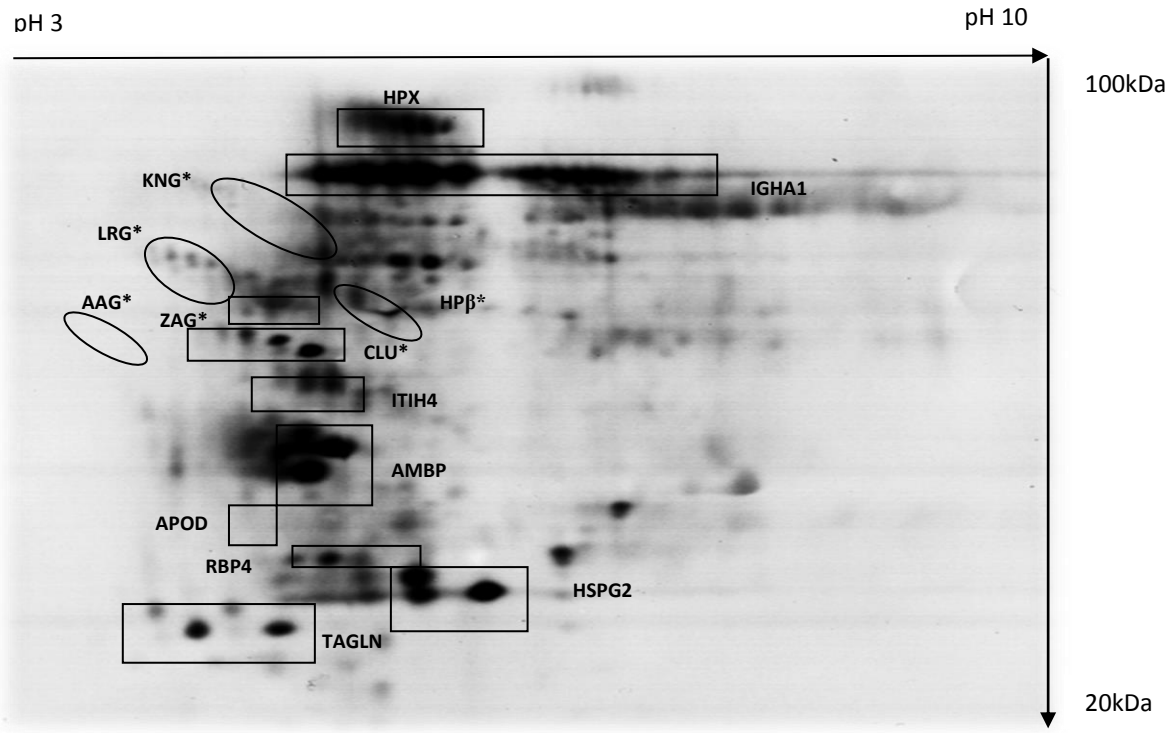
**A: Control**




**B: ECa**



**C: OCa**


pH 3
pH 10
100kDa
20kDa
HPX
KNG*
IGHA1
LRG*
AAG*
ZAG*
HPβ*
CLU*
ITIH4
AMBP
APOD
RBP4
HSPG2
TAGLN

**D: CCa**

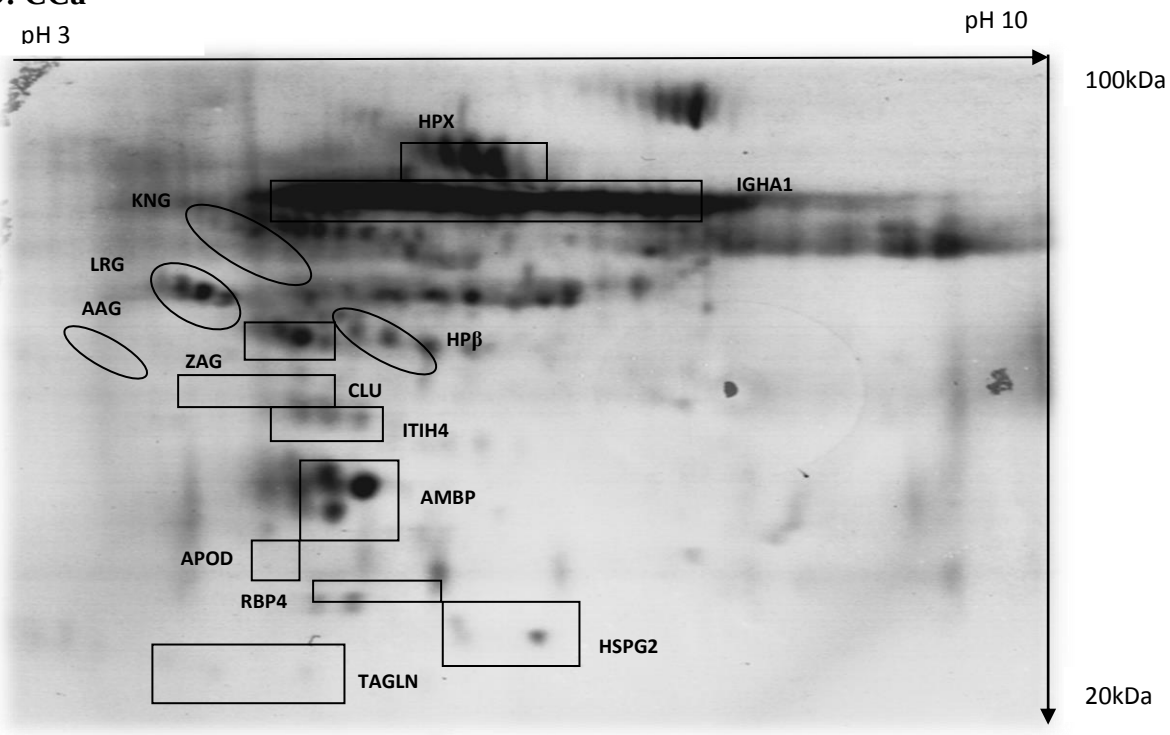

pH 3
pH 10
100kDa
20kDa
HPX
IGHA1
KNG
LRG
AAG
HPβ
ZAG
CLU
ITIH4
AMBP
APOD
RBP4
HSPG2
TAGLN

Table 3.4 shows the relative level of CMB lectin detected urinary glycoproteins in the three groups of cancer patients studied.

When image analysis was performed on the CMB blots from patients with CCa as well as those of the controls, there was no significant difference between the level of all detected urinary glycoproteins. Figure 3.8 shows the mean percentage of volume contribution of CMB lectin detected urinary glycoproteins from the controls and patients with ECa, OCa and CCa.

**Table 3.4: The relative level of *N*-glycosylated urinary glycoprotein**

| ***N*-glycosylated urinary glycoprotein** | **Fold of changes** | | |
|---|---|---|---|
| | **ECa** | **OCa** | **CCa** |
| KNG | n.s. | 3.29 | n.s. |
| LRG | n.s. | +6.07 | n.s. |
| HPβ | n.s. | +63.33 | n.s. |
| ZAG | +2.79 | +4.05 | n.s. |
| AAG | +31.8 | +58.23 | n.s. |
| CLU | n.s. | +5.20 | n.s. |

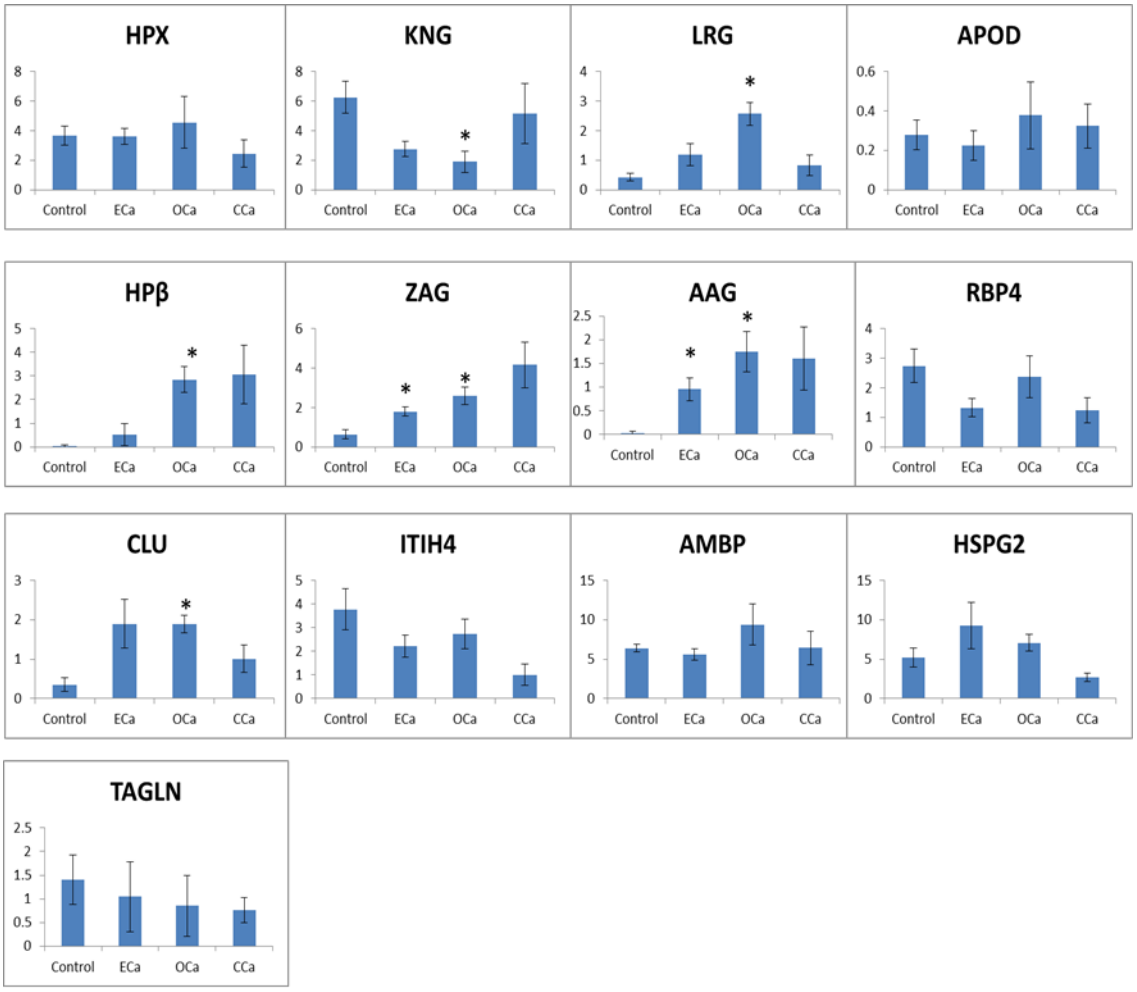
Fold changes were relative to that of control
(-) decrease in level; (+) increase in level
n.s.: not significant

## 3.7 Identification of urinary proteins using mass spectrometry

Mass spectrometry analysis was performed using the 4800 Plus MALDI TOF/TOF analyzer (Applied Biosystem/MDS Sciex, Toronto, Canada). The experimental peptide masses were then submitted to the MASCOT database to search

**Figure 3.8: Mean percentage of volume contribution of CMB lectin detected urinary glycoproteins from controls and patients with ECa, OCa and CCa**

Percentage of volume contribution of each urinary glycoproteins for control (n=11), ECa (n=7), OCa (n=9) and CCa (n=8) was calculated using Image Master 2D Platinum Software 7.0. The value was expressed as mean of the percentages of volume contribution ± S.E.M. Student *t*-test was used to analyze the significant difference between the control and the patients. False discovery rate was performed using the method of Benjamini and Hochberg. Proteins marked with * were significantly different compared to that of control. X-axis represents the sample group and Y-axis represents the percentage of volume contribution.


HPX
KNG
LRG
APOD
HPβ
ZAG
AAG
RBP4
CLU
ITIH4
AMBP
HSPG2
TAGLN
Control
ECa
OCa
CCa

against known theoretical peptide masses of proteins from *Homo sapiens*. Database search parameters were set as: the enzyme trypsin was used; one missed cleavage was allowed in trypsin digestion; variable modification included were carbamidomethylation of cysteine and oxidation of methionine; the mass tolerance for MS precursor ion and MS/MS fragment ion were 50 ppm and 0.1 Da, respectively; and only monoisotopic masses were included in the search. The identification of protein clusters that consistently appeared in 2-DE silver-stained profiles was done by subjecting the spot clusters of urine proteins to in-gel digestion, followed by the analysis with the mass spectrometry and database search. The identities of these protein clusters were confirmed as KNG, AAG, ZAG, CD59, AMBP, IGHG3 and KAC.

For protein clusters that did not appear by silver staining but only detected using lectins, on membrane digestion (Section 2.2.7.2) was performed prior to the analysis by mass spectrometry. NEB and CLU were successfully identified using this technique. Table 3.5 shows the mass spectrometric identification data of the urinary proteins that were resolved by 2-DE and using silver staining, CGB lectin and CMB lectin for detection. The HP cluster spots detected on the CMB blot was actually the β chain of protein. The mass spectrometry analysis generated peptides sequences with high confidence which spanned within the β chain region of HP. Table 3.6 shows the list of matched peptide sequences with high confidence for HP. On the other hand, NEB was suggested to be a truncated fragment of its native form due to the large discrepancy presented between its experimental mass and theoretical mass.

**Table 3.5: Mass spectrometric identification of 2-DE resolved urinary proteins**

| Protein Entry Name[+] | Protein name | Accession number[#] | Nominal mass (kDa)/p*I* | MOWSE protein score | Sequence coverage (%) |
|---|---|---|---|---|---|
| **NEB*** | Nebulin | P20929 | 775/9.11 | 64 | 1[#] |
| **CLU*** | Clusterin | P10909 | 52/5.89 | 25 | 11 |
| **KNG** | Kininogen | P01042 | 71/6.34 | 68 | 7 |
| **IGHG3** | Ig γ3 chain C region | P01860 | 41/8.46 | 16 | 3 |
| **AAG** | α1-acid glycoprotein | P19652 | 23/5.03 | 241 | 16 |
| **ZAG** | Zinc α2 glycoprotein | P25311 | 33/5.57 | 134 | 20 |
| **AMBP** | Protein AMBP | P02760 | 39/5.95 | 50 | 12 |
| **IGKC** | Ig λ1 chain C region | P01834 | 11/5.58 | 141 | 32 |
| **CD59** | CD59 glycoprotein | P13987 | 14/6.02 | 121 | 18 |
| **HP** | Haptoglobin | P00738 | 45/6.13 | 313 | 18 |
| **LRG** | Leucine-rich α2 glycoprotein | P02750 | 38/6.45 | 149 | 10 |
| **ITIH4** | Inter alpha trypsin inhibitor heavy chain H4 | Q14624 | 103/6.51 | 336 | 14 |

*The protein was prepared by on-membrane digestion

[#]Relative to the sequence of native nebulin

[+]Protein entry names are from the UniProtKB database http://www.uniprot.org assessed on 22 Feb 2012

Accession numbers are from the Mascot search engine

**Table 3.6: The list of matched peptide sequences with high confidence for haptoglobin**

| Peptide sequence | Amino acid |
|---|---|
| R.VGYVSGWGR.N | 278-286 |
| K.YVMLPVADQDQCIR.H | 298-311 |
| K.SPVGVQPILNEHTFCAGMSK.Y | 326-345 |
| K.YQEDTCYGDAGSAFAVHDLEEDTWYATGILSFDK.S | 346-379 |

Sequences of peptides were checked using the Swiss-Prot database.

## 3.8 LC-MS/MS analysis of lectin affinity captured urinary glycoproteins

### 3.8.1 CGB lectin affinity chromatography

To analyse native *O*-glycosylated proteins, 400 μg of pooled urinary protein samples of eleven control subjects, seven ECa patients, nine OCa patients and eight CCa patients were respectively subjected to the CGB lectin affinity chromatography as described in Section 2.2.8.2. The columns were washed extensively with PBS at pH 7.2 to remove non-binding proteins and other substances. Fractions of 1 ml were collected and monitored at 280 nm until the absorbance reached the baseline. Bound *O*-glycosylated urinary proteins were eluted with 0.1 M melibiose. Figure 3.9 shows the elution profiles of the bound fractions of CGB lectin affinity chromatography for controls and the three groups of cancer patients.

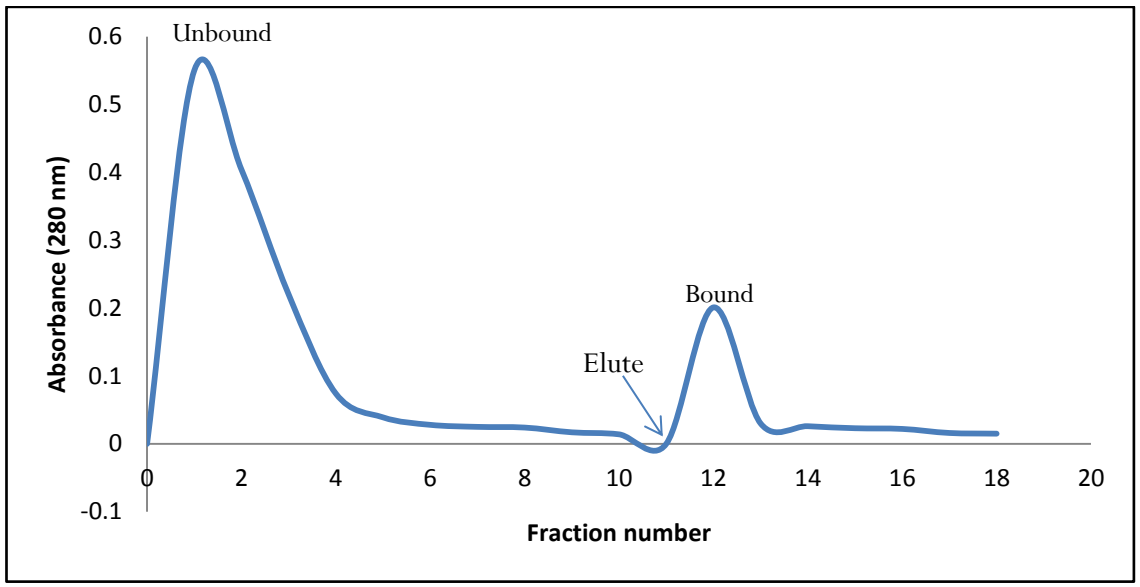
### 3.8.2 CMB lectin affinity chromatography

Native *N*-glycosylated urinary proteins were also analysed using the CMB lectin affinity chromatography as described in Section 2.2.8.2. Similar to the CGB chromatography fractionation procedure, pooled urine samples of the controls, ECa patients, OCa patients and CCa patients were respectively loaded into separate CMB affinity columns and washed extensively with PBS. Fractions of 1 ml were collected and monitored at 280 nm until the absorbance level reached the baseline. The bound *N*-glycosylated urinary proteins were eluted with 0.1 M methyl-α-D-mannopyranoside. Figure 3.10 shows the elution profiles of the bound fractions of CMB lectin affinity chromatography for the controls and three groups of cancer patients.

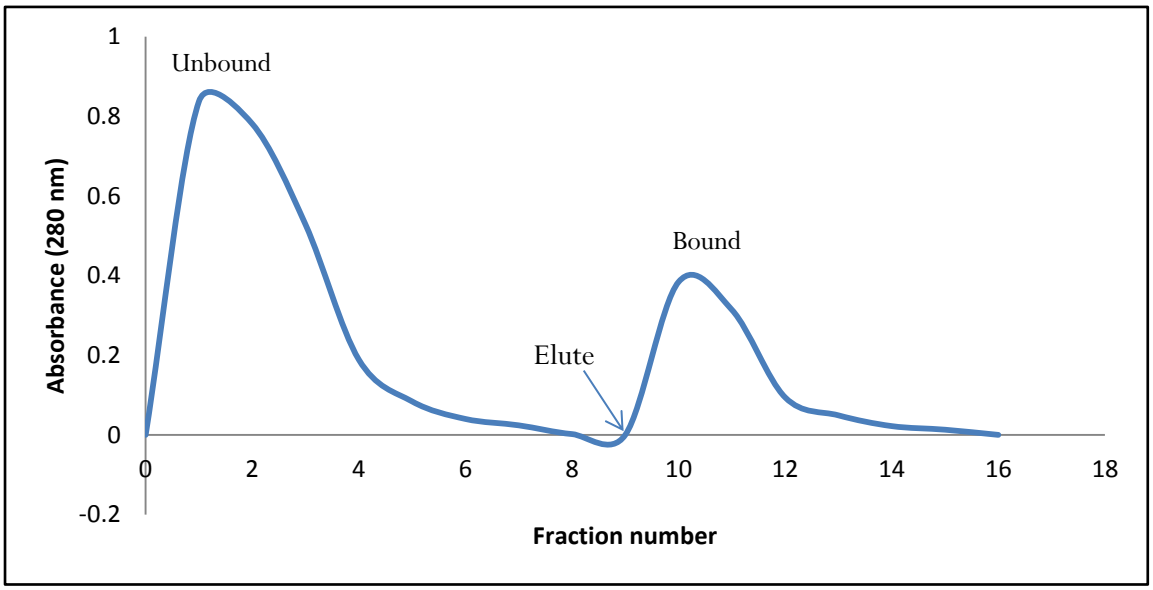
**Figure 3.9: Elution profiles of bound fractions of CGB lectin affinity chromatography**

Each pooled urine samples were subjected to the CGB lectin affinity column chromatography. Unbound fractions were washed out with PBS, pH 7.2 until the absorbance reached the baseline. Bound glycoproteins were eluted out with 0.1 M melibiose with gravitational flow. Four different elution profiles of controls (panel A), and patients with ECa (panel B), OCa (panel C) and CCa (panel D) were generated. Arrow indicates the start of sugar elution.

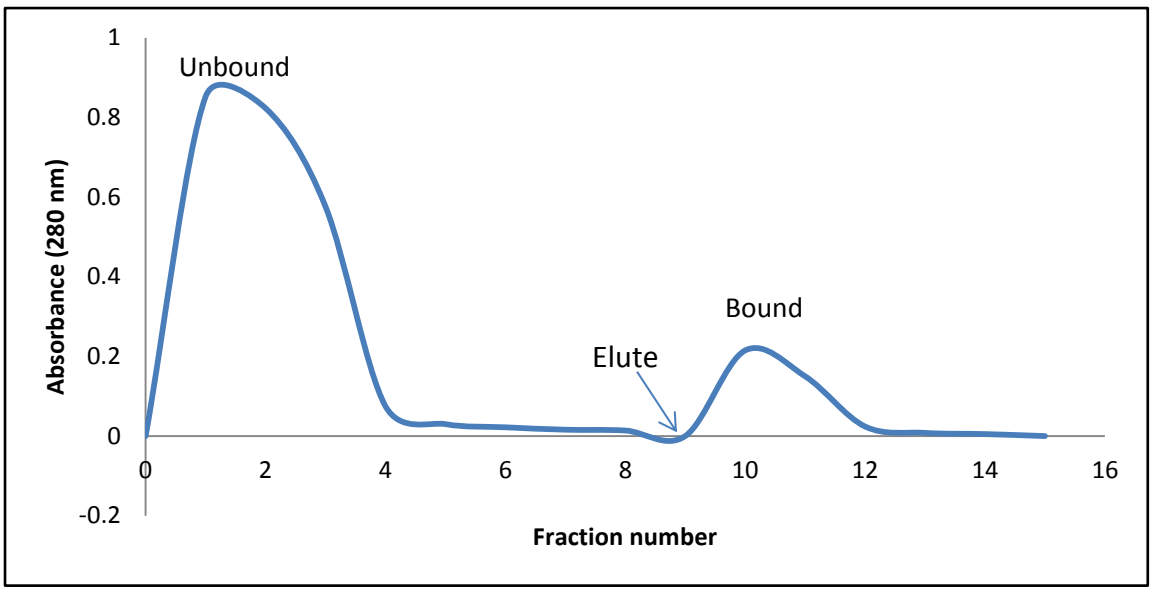
**A: Control**




**B: ECa**



**C: OCa**


Unbound
Bound
Elute
Absorbance (280 nm)
Fraction number
1
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16

**D: CCa**

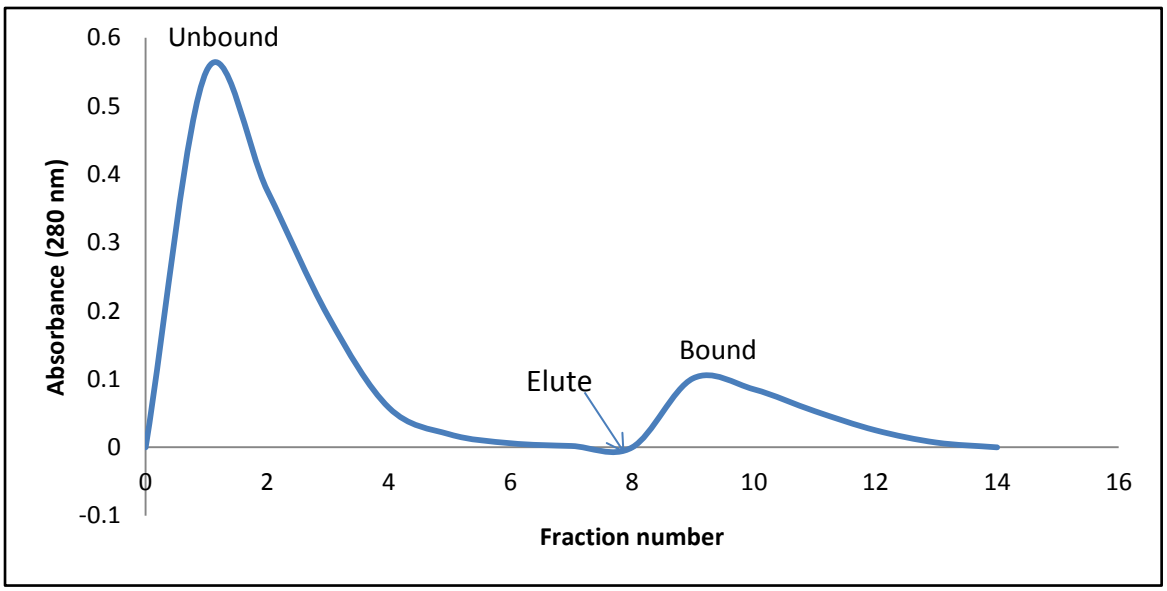
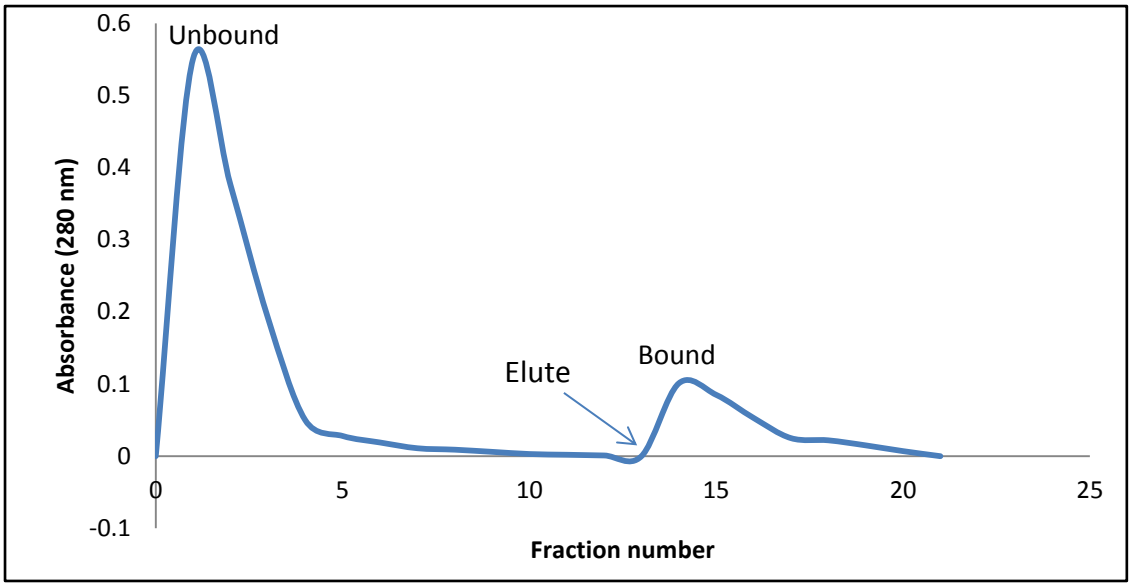

Unbound
Bound
Elute
Absorbance (280 nm)
Fraction number
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
2
4
6
8
10
12
14
16

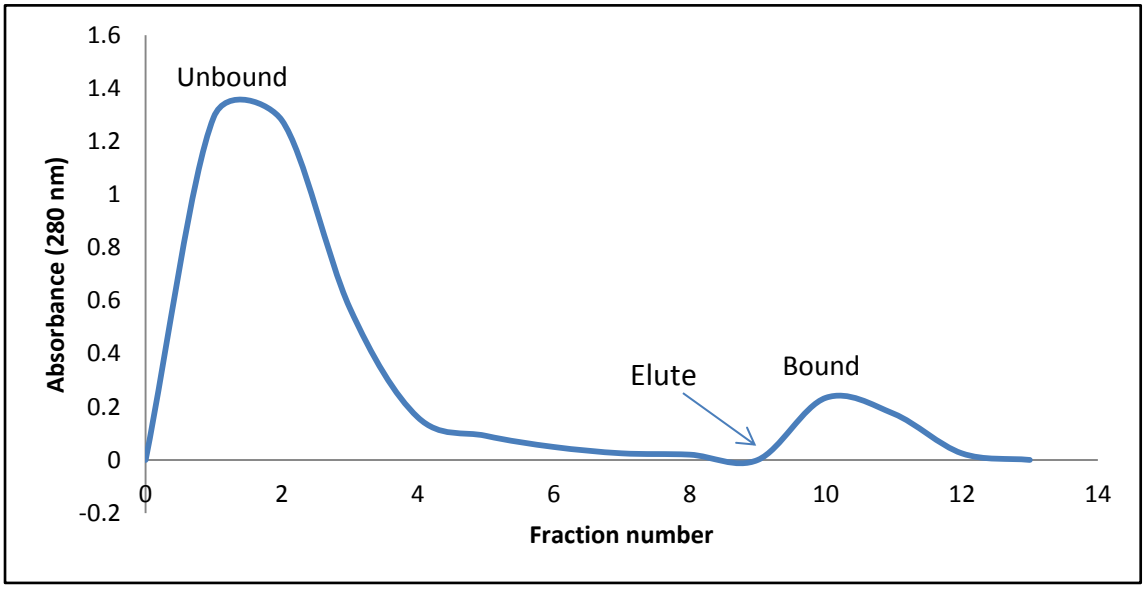
**Figure 3.10: Elution profiles of bound fractions of CMB lectin affinity chromatography**

Each pooled urine samples were loaded into the CMB lectin affinity column and 1 ml of fractions were collected. Unbound fractions were washed out with PBS, pH 7.2 until absorbance fell to the baseline. Elution of bound glycoproteins were carried out using 0.1 M methyl-α-D-mannopyranoside. Four different elution profiles of control (panel A), and patients with ECa (panel B), OCa (panel C) and CCa (panel D) were generated. Arrow indicates the start of sugar elution.

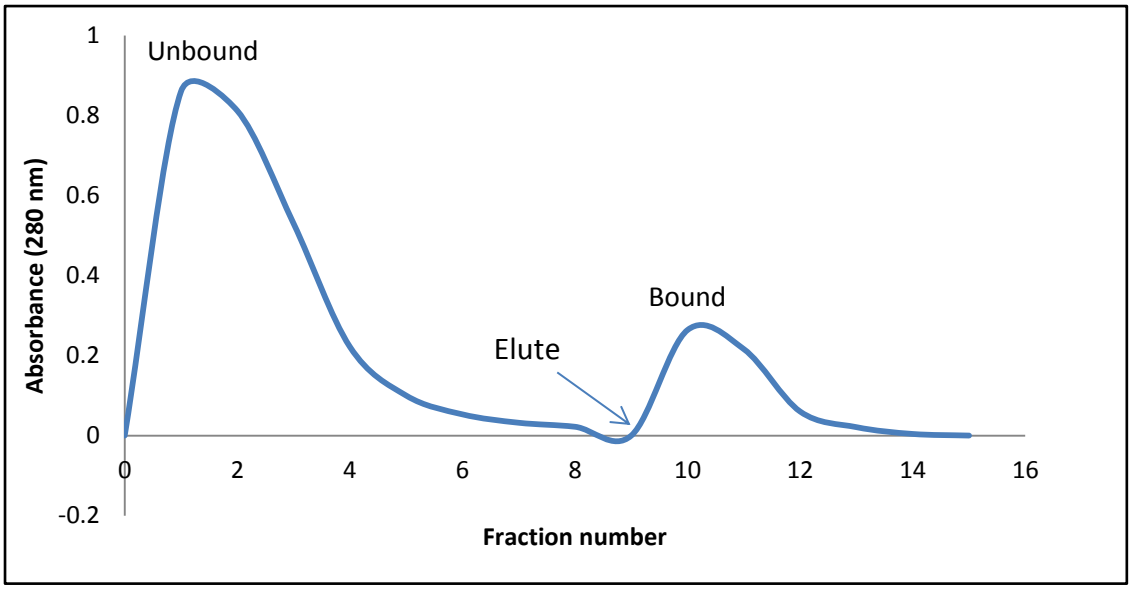
**A: Control**



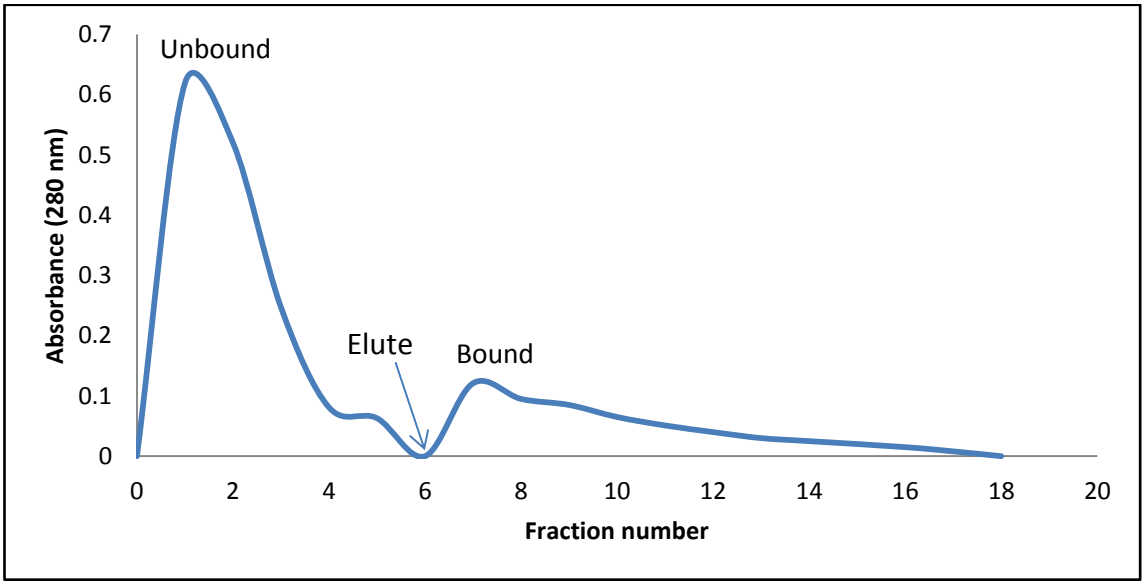

**B: ECa**



**C: OCa**


Unbound
Bound
Elute
Absorbance (280 nm)
Fraction number
1
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16

**D: CCa**


Unbound
Elute
Bound
Absorbance (280 nm)
Fraction number
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20

### 3.8.3 Profiling of native *O*-glycosylated urinary proteins captured by CGB lectin

Profiling of native *O*-glycosylated urinary proteins was performed by subjecting the bound protein fractions of each groups of the immobilized CGB-lectin affinity chromatography to LC-MS/MS analysis separately. In this study, a total of 36 proteins were identified with high confidence levels. Out of the 36 identified proteins, ten were annotated as *O*-glycosylated by the Swiss-Prot database. Another 26 were annotated as potential *O*-glycosylated urinary proteins as predicted by the NetOGlyc 3.1 server. The *O*-glycosylated urinary proteins were categorized based on their subcellular localization. Based on the Swiss-Prot database, there were 11 secreted proteins, 15 membrane proteins, four nuclear proteins, four cytoplasmic proteins and two unknown proteins. Table 3.7 summarizes all of the identified *O*-glycosylated proteins or potential *O*-glycosylated proteins that were captured by the CGB lectin affinity column chromatography.

**Table 3.7: List of *O*-glycosylated/potentially *O*-glycosylated proteins isolated using CGB lectin**

| Protein name | Accession number | Subcellular location | Glycan# |
|---|---|---|---|
| Protein AMBP | P02760 | Secreted | *O*-linked |
| ATP synthase subunit beta, mitochondrial | P24539 | Membrane | Potential *O*-linked |
| Serotransferrin | P02787 | Secreted | *O*-linked |
| Transmembrane protein 110 | Q86TLZ | Membrane | Potential *O*-linked |
| Phosphoinositide-3 kinase interacting protein 1 | Q96FE7 | Membrane | *O*-linked |
| Ribonuclease pancreatic | P07998 | Secreted | Potential *O*-linked |
| Prostaglandin-H2-D isomerase | P41222 | Secreted | Potential *O*-linked |
| Membrane bound transcription factor site-1 protease | Q14703 | Membrane | Potential *O*-linked |
| Homeobox protein engrailed 2 | P19622 | Nucleus | Potential *O*-linked |
| Actin, cytoplasmic 1 | P60709 | Cytoplasmic | Potential *O*-linked |
| Neurofilament medium polypeptide | P07197 | Cytoplasmic | *O*-linked |

Continue on next page.

Continuation from previous page.

| | | | |
|---|---|---|---|
| CD55 decay-accelerating factor splicing variant 4 | Q14UF3 | Membrane | Potential *O*-linked |
| Leucine rich alpha 2 glycoprotein | P02750 | Secreted | *O*-linked |
| Spondin-2 | Q9BUD6 | Secreted | Potential *O*-linked |
| Protein RRP5 homolog | Q14690 | Nucleus | Potential O-linked |
| CD44 antigen | P16070 | Membrane | *O*-linked |
| Phosphatidylinositol-3,4,5-triphosphate 5-phosphatase-1 | Q92835 | Membrane | Potential *O*-linked |
| Solute carrier family 12 member 6 | Q9UHW9 | Membrane | Potential *O*-linked |
| Mucin-5B | Q9HC84 | Secreted | *O*-linked |
| Major facilitator superfamily domain containing protein 10 | Q14728 | Membrane | Potential *O*-linked |
| PDZ and LIM domain protein 5 | Q96HC4 | Membrane | Potential *O*-linked |
| Leukocyte associated immunoglobulin-like receptor 1 | Q6GTX8 | Membrane | Potential *O*-linked |
| Tumor necrosis receptor superfamily member 16 | P08138 | Membrane | Potential *O*-linked |
| Lipocalin-1 | P31025 | Secreted | Potential *O*-linked |
| Amyloid beta A4 | P05067 | Membrane | *O*-linked |
| Nebulin | P20929 | Cytoplasmic | Potential *O*-linked |
| Transcription factor HES-2 | Q9Y543 | Nucleus | Potential *O*-linked |
| Rho GTPase-activating protein 12 | P20936 | Cytoplasmic | Potential *O*-linked |
| Solute carrier family 13 member 3 | Q8WWT9 | Membrane | Potential *O*-linked |
| Sodium/bile acid contransporter | Q14973 | Membrane | Potential *O*-linked |
| Bromodomain-containing protein 3 | Q15059 | Nucleus | Potential *O*-linked |
| Inter alpha trypsin inhibitor heavy chain H4 | Q14624 | Secreted | *O*-linked |
| Ig gamma-1 chain C region | P01857 | Secreted | Potential *O*-linked |
| Ig gamma-2 chain C region | P01859 | Secreted | Potential *O*-linked |
| Ig alpha-1 chain C region | P01876 | Unknown | *O*-linked |
| Ig lambda-1 chain C region | P0CG04 | Unknown | Potential *O*-linked |

# Type of glycan was annotated by UniProt; glycosylation potential was determined by NetOGlyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc).

### 3.8.4 Profiling of native *N*-glycosylated urinary proteins captured by CMB lectin

Profiling of native *N*-glycosylated proteins of controls and patients with ECa, OCa and CCa were performed by subjecting pooled urine samples from controls and patients to immobilized CMB lectin affinity chromatography separately and analysis of their bound fractions using LC-MS/MS. Out of the 46 identified proteins with high confidence level, 28 were annotated as proteins that are *N*-glycosylated by the Swiss-

Prot database. Another 18 were annotated as potential *N*-glycosylated proteins, as predicted by the NetNGlyc 1.0. Based on the Swiss-Prot database, 17 were secreted proteins, 14 were membrane proteins, 5 were nuclear proteins, 5 were cytoplasmic proteins, 1 was lysosomal protein, 1 was endoplasmic reticulum protein, 1 was mitochondrion protein and 2 were unknown proteins. Table 3.8 summarizes all of the identified *N*-linked glycoprotein or potential *N*-linked glycoproteins captured by CMB-lectin affinity column chromatography.

**Table 3.8: List of *N*-glycosylated/potentially *N*-glycosylated proteins isolated using CMB lectin**

| Protein name | Accession number | Subcellular location | Glycan# |
|---|---|---|---|
| Protein AMBP | P02760 | Secreted | *N*-linked |
| Zinc alpha-2 glycoprotein | P02768 | Secreted | *N*-linked |
| Serotransferrin | P02787 | Secreted | *N*-linked |
| Uromodulin | P07911 | Secreted | *N*-linked |
| Transmembrane protein 110 | Q86TL2 | Membrane | Potential *N*-linked |
| Transmembrane protein 43 | Q9BTV4 | Nucleus | Potential *N*-linked |
| Phosphoinositide-3-kinase interacting protein 1 | Q96FE7 | Membrane | *N*-linked |
| Alpha 1-acid glycoprotein | P19652 | Secreted | *N*-linked |
| Ig kappa chain C region | P01834 | Unknown | Not *N*-linked |
| LRP11 protein | Q96A73 | Unknown | Not *N*-linked |
| E3 ubiquitin-protein ligase RNF149 | Q8NC42 | Membrane | *N*-linked |
| Ribonuclease pancreatic | P07998 | Secreted | *N*-linked |
| Prostaglandin-H2 D-isomerase | P41222 | Secreted | *N*-linked |
| Membrane bound transcription factor site-1 protease | Q14703 | Membrane | *N*-linked |
| RecQ-mediated genome instability protein 1 | Q9H9A7 | Nucleus | Potential *N*-linked |
| G protein coupled receptor 87 | Q9BY21 | Membrane | *N*-linked |
| Homeobox protein engrailed-2 | P19622 | Nucleus | Potential *N*-linked |
| Ig alpha-1 chain C region | P01876 | Unknown | *N*-linked |
| Non-secretory ribonuclease | P10153 | Cytoplasmic | *N*-linked |
| CD44 antigen | P16070 | Membrane | *N*-linked |
| Galectin-3 binding protein | P17931 | Secreted | Potential *N*-linked |
| Grainyhead-like protein 1 homolog | Q9NZ15 | Nucleus | Potential *N*-linked |
| WAP four-disulfide core domain protein 2 | Q14508 | Secreted | *N*-linked |

Continue on next page.

Continuation from previous page.

| | | | |
|---|---|---|---|
| CD59 glycoprotein | P13987 | Secreted | *N*-linked |
| Kininogen | P01042 | Secreted | *N*-linked |
| Hemopexin | P02790 | Secreted | *N*-linked |
| Granulins | P28799 | Secreted | *N*-linked |
| N-acetylglucosamine 6-sulfatase | P15586 | Lysosome | *N*-linked |
| Lysosome associated membrane glycoprotein 2 | P13473 | Membrane | *N*-linked |
| Coiled- coil domain containing protein 125 | Q86Z20 | Cytoplasmic | Potential *N*-linked |
| ATrich interactive domain containing protein 4B | Q4LE39 | Cytoplasmic | Potential *N*-linked |
| Basement membrane specific heparin sulfate proteoglycan core protein | P98160 | Secreted | *N*-linked |
| Symplekin | Q92797 | Cytoplasmic | Potential *N*-linked |
| Collagen-3 (VI) chain | Q01955 | Secreted | *N*-linked |
| MAGUK p55 subfamily member 7 | Q5T2T1 | Nucleus | Potential *N*-linked |
| Ran binding protein 17 | Q9H2T7 | Membrane | *N*-linked |
| Propionyl-CoA carboxylase alpha chain, mitochondrion | P05165 | Mitochondrion | Potential *N*-linked |
| PGAP2-interacting protein | Q9H720 | Membrane | *N*-linked |
| Prostaglandin F2 receptor negative regulator | P41222 | Secreted | *N*-linked |
| Growth hormone receptor | P10912 | Membrane | *N*-linked |
| Serine/threonine kinase 11 interacting protein | Q8N1F8 | Cytoplasmic | Potential *N*-linked |
| Probable phospholipid transporting ATPase IIA | O75110 | Membrane | Potential *N*-linked |
| Solute carrier family 26 member 9 | Q8BU91 | Membrane | Potential *N*-linked |
| ATP binding cassette subfamily B member 10, mitochondrion | Q9NRK6 | Membrane | Potential *N*-linked |
| Rab II family interacting protein 1 | Q6WKZ4 | Membrane | Potential *N*-linked |
| Protein disulfide-isomerase A2 | Q13087 | Endoplasmic reticulum | *N*-linked |

# Type of glycan was annotated by UniProt; glycosylation potential was determined by NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/)

### 3.8.5 Protein quantification using spectral counting

To validate the altered levels of the urinary proteins that were shown by 2-DE and Western blot-lectin analyses, proteins that were subjected to LC-MS/MS analysis (i.e., KNG, AAG, ZAG, CD59, LRG and NEB) were further analyzed using spectral counting approach. The technique involves relative protein quantification based on the normalized spectral abundance factor (NSAF) values that were generated from the

number of identified MS/MS spectra of a protein and its peptide length. Among the proteins that were analysed, only LRG and NEB were *O*-glycosylated. Analysis of the NSAF values of LRG appears to indicate that its concentration was higher in patients with OCa compared to the controls, where else the concentrations of urinary LRG of patients with ECa and CCa were comparable to the controls. On the other hand, the NEB peptide spectrum with the *de novo* sequence K.AYELQSDNVYKADLEWLRGIGWMPNDSVSVNHA.K (amino positions 4103-4136) was apparently present only in the pooled urine samples of the controls but not detected in the pooled urine samples of patients with ECa.

Based on the spectral count of bound fractions from CMB lectin affinity chromatography, the concentrations of ZAG and AAG were found to be higher in the pooled urine samples from patients with ECa. Apart from that, the concentration of KNG was lower in bound fractions of CMB-lectin affinity columns from patients with OCa. Eleven spectrums for HP were solely present in the urine samples from patients with OCa when their pooled bound fractions of CMB lectin affinity columns were analyzed. Similar to NEB, two CD59 peptide spectrums with the de novo sequence K.FEHCNFNDVTTR.L (amino position 67-78) appeared only in the pooled urine samples of controls but not in the pooled urine samples of all cohorts of patients. Table 3.9 summarizes the number of spectra and the ratio of the altered proteins which were previously detected to be significantly different in 2-DE and Western blot-lectin analysis.

**Table 3.9: The number of spectra of differentially expressed urinary proteins obtained from 2-DE silver-stained analysis and Western blot-lectin analysis**

| | **Number of spectra of *O*-glycosylated urinary proteins** | | | | | **Number of spectra of *N*-glycosylated urinary proteins** | | |
|---|---|---|---|---|---|---|---|---|
| Protein | Control | ECa | OCa | CCa | Control | ECa | OCa | CCa |
| KNG | - | - | - | - | 4 | 3 | 1 | 3 |
| AAG | - | - | - | - | 6 | 11 | 4 | 5 |
| ZAG | - | - | - | - | 19 | 38 | 13 | 16 |
| LRG | 8 | 9 | 29 | 10 | - | - | - | - |
| NEB | 1 | - | 1 | 1 | - | - | - | - |
| CD59 | - | - | - | - | 2 | - | - | - |
| HP | - | - | - | - | - | - | 11 | - |

-: not detected

## 3.9 Competitive ELISA

Since CLU was not detected in both CGB and CMB lectins bound fractions, competitive ELISA was performed using monoclonal antisera against CLU to validate the altered levels of the urinary proteins that were detected in the patients with OCa ($n$ = 9) compared to the control subjects ($n$ = 11). Figure 3.11 demonstrates the mean percentage of inhibition of CLU that was tested in respective groups of samples. The expression of CLU was significantly enhanced in almost all of the urine samples of the OCa patients compared to the controls.

## 3.10 SELDI-TOF-MS analysis of urinary glycopeptides

In this study, a preliminary investigation was performed to determine if the CMB and CGB lectins may be used to discriminate digested gycopeptides that are present in the urine of patients with ECa, CCa and OCa, compared to the controls using the SELDI-TOF-MS. The SELDI-TOF-MS is restricted to the analysis of low molecular weight peptide. Thus, subsequent analysis of urinary proteins from controls

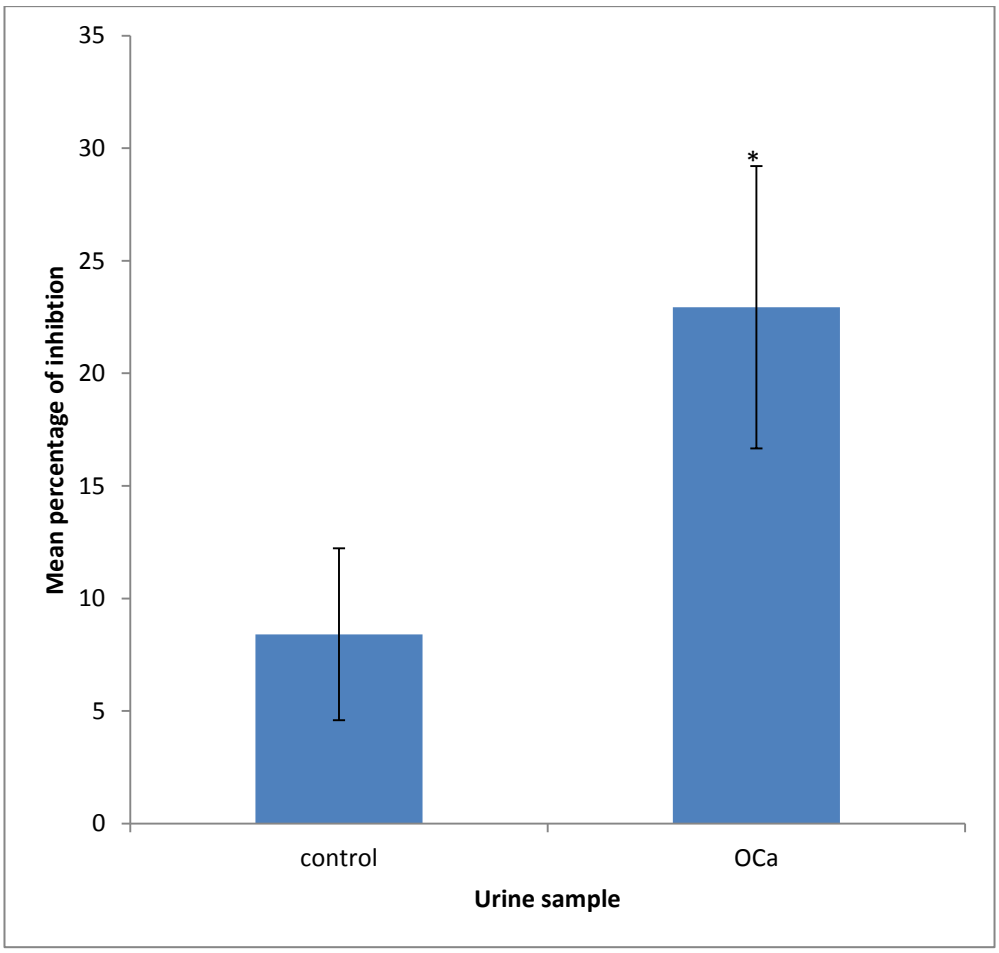



**Figure 3.11: Estimation of CLU using competitive ELISA**

Urinary proteins from controls and patients with OCa were incubated with anti-CLU monoclonal antibody in microplate wells. The amount of CLU present in the samples was proportional to the percentage of inhibition. * denotes the percentage of inhibition was statistically significant with $p < 0.05$.

and patients, as described in the following section, initially involved the preparation of peptides from the urinary protein samples using trypsin digestion. This was followed by selective capture of the *N*- or *O*-glycopeptides using the respective CMB and CGB lectins that were immobilised onto the PS10 chips (Section 3.10.1). Profiling of the glycopeptides was performed by detection and alignment of the peaks and analysis of the data generated using a decision tree algorithm (Section 3.10.2-3). Potential peaks that can truly discriminate between each groups of samples were finally identified (Section 3.10.4).

### 3.10.1 Detection of peptide peaks

From the SELDI spectra of the CMB and CGB lectin captured urinary glycopeptide profiles, a total of 54 and 24 raw peaks were detected between *m/z* 1000 and 10000, respectively. Figure 3.12 shows the representative spectra views for the CMB and CGB lectin captured urinary glycopeptides of controls and patients with ECa CCa and OCa, respectively.

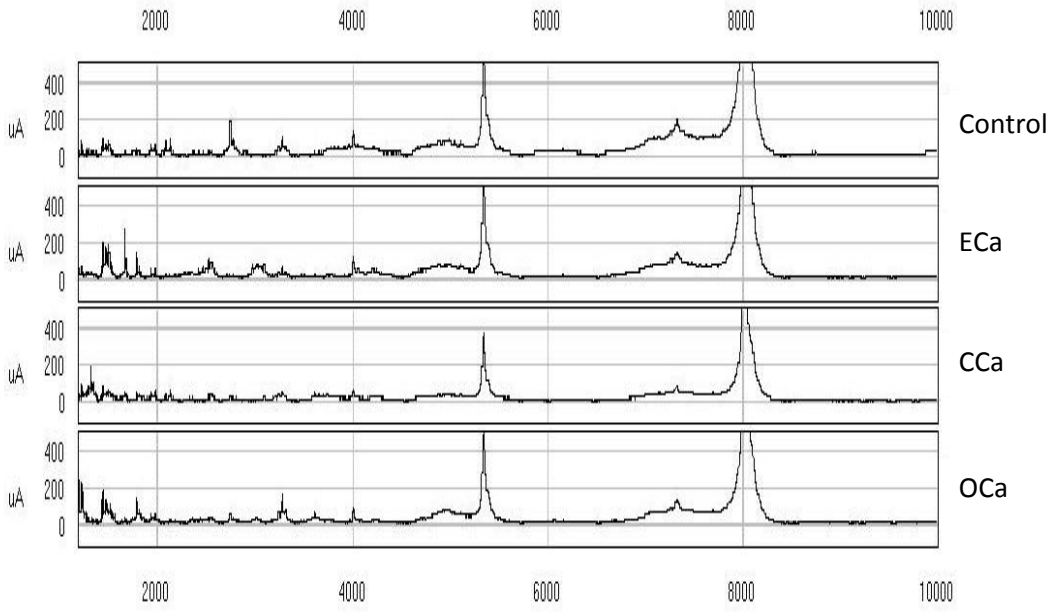
### 3.10.2 Identification of potential biomarker pattern and construction of Diagnostic Model of CMB lectin captured urinary glycopeptide profile

When comparison between the CMB lectin captured urinary glycopeptides of controls and patients with ECa, OCa and CCa were performed, two peaks (*m/z* 1201 and 1449) were selected by the Biomarker Pattern Software 5.0 to set up the decision tree for group discrimination. Figure 3.13 shows the tree structure and the sample distribution for these peaks in the CMB lectin captured urinary glycopeptide profile. At node 1, *m/z* 1201 was used to classify the samples, where samples having the peak

**Figure 3.12: Representative spectral view of CMB and CGB lectin captured urinary glycopeptides**

Urinary peptides were incubated on lectins immobilized PS10 chips and subjected to SELDI-TOF analysis. *N*-linked or *O*-linked glycopeptides captured by CMB (panel A) or CGB lectin (panel B), respectively were ionized by laser energy and measured as *m/z* (y-axis) by the ion chamber of the instrument. The detected urinary glycopeptides are expressed as a series of peaks with their own intensity, μA (y-axis).

**A**


2000
4000
6000
8000
10000
uA
400
200
0
Control
ECa
CCa
OCa

**B**

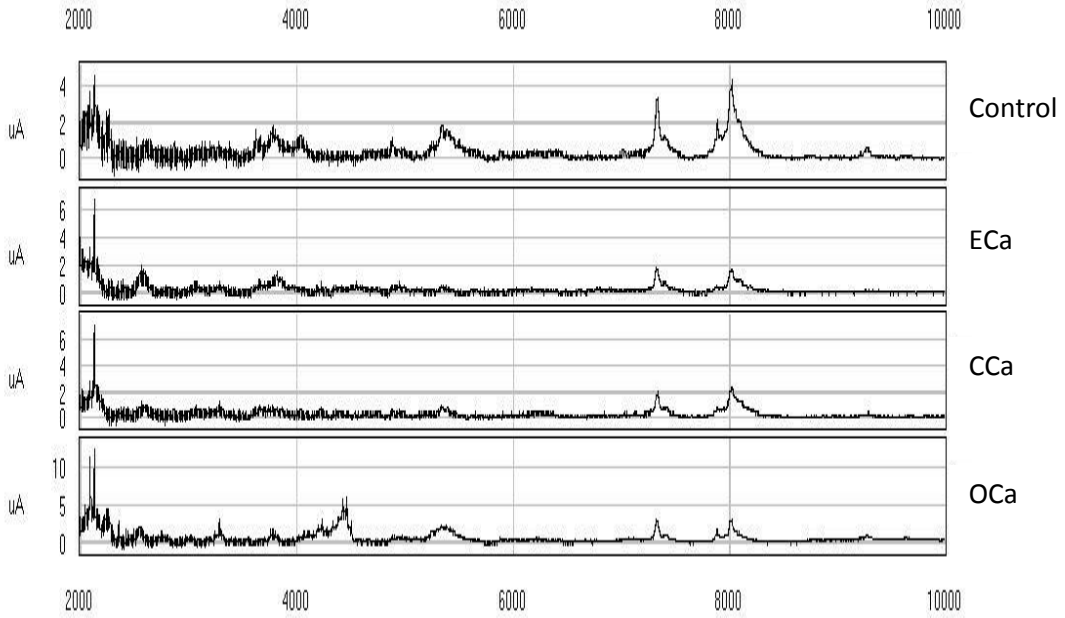
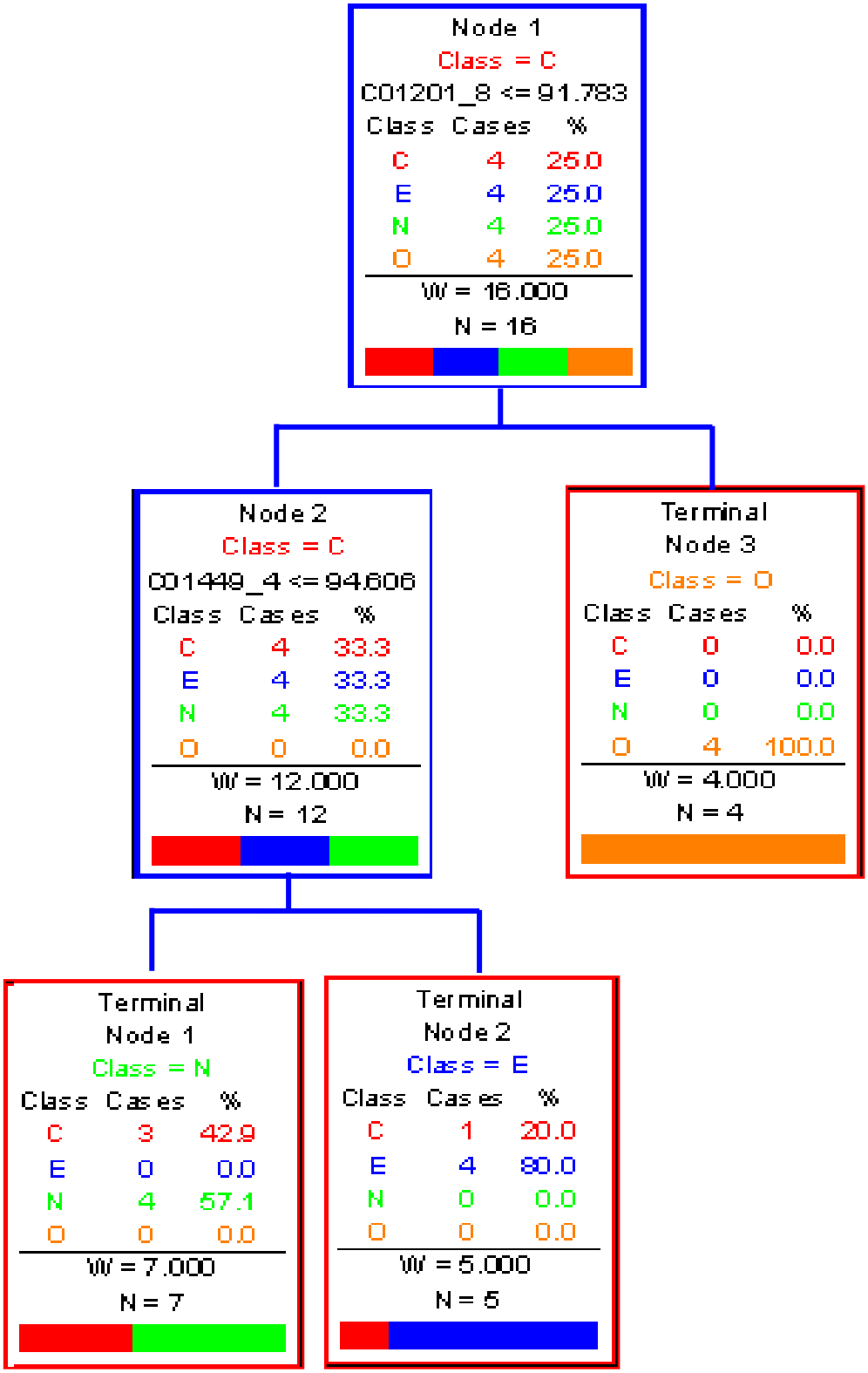

2000
4000
6000
8000
10000
uA
4
2
0
6
10
5
Control
ECa
CCa
OCa

**Figure 3.13: Tree structure profiles and samples distribution of CMB lectin captured urinary glycopeptides**

Peaks with *m/z* of 1201 and 1449 were chosen to set up the decision tree. C, E, N and O refer to CCa, ECa, control and OCa accordingly. The root node and the terminal node were indicated as blue and red colours, respectively.


Node 1
Class = C
C01201_8 <= 91.783
Class Cases %
C 4 25.0
E 4 25.0
N 4 25.0
O 4 25.0
W = 16.000
N = 16
Node 2
Class = C
C01449_4 <= 94.606
Class Cases %
C 4 33.3
E 4 33.3
N 4 33.3
O 0 0.0
W = 12.000
N = 12
Terminal
Node 3
Class = O
Class Cases %
C 0 0.0
E 0 0.0
N 0 0.0
O 4 100.0
W = 4.000
N = 4
Terminal
Node 1
Class = N
Class Cases %
C 3 42.9
E 0 0.0
N 4 57.1
O 0 0.0
W = 7.000
N = 7
Terminal
Node 2
Class = E
Class Cases %
C 1 20.0
E 4 80.0
N 0 0.0
O 0 0.0
W = 5.000
N = 5

intensity higher than 91.783 μA went to Terminal node 3. All four samples of OCa went into this node. The rest of the samples entered Node 2, which had four samples each for the remaining groups. At node 2, samples having *m/z* 1449 with peak intensity higher than 94.606 μA went to Terminal node 2. One sample of CCa and four samples of ECa fell into this node. The remaining samples entered Terminal node 1 with three samples of CCa and four samples of controls.

### 3.10.3 Identification of potential biomarker pattern and construction of Diagnostic Model of CGB lectin captured urinary glycopeptide profile

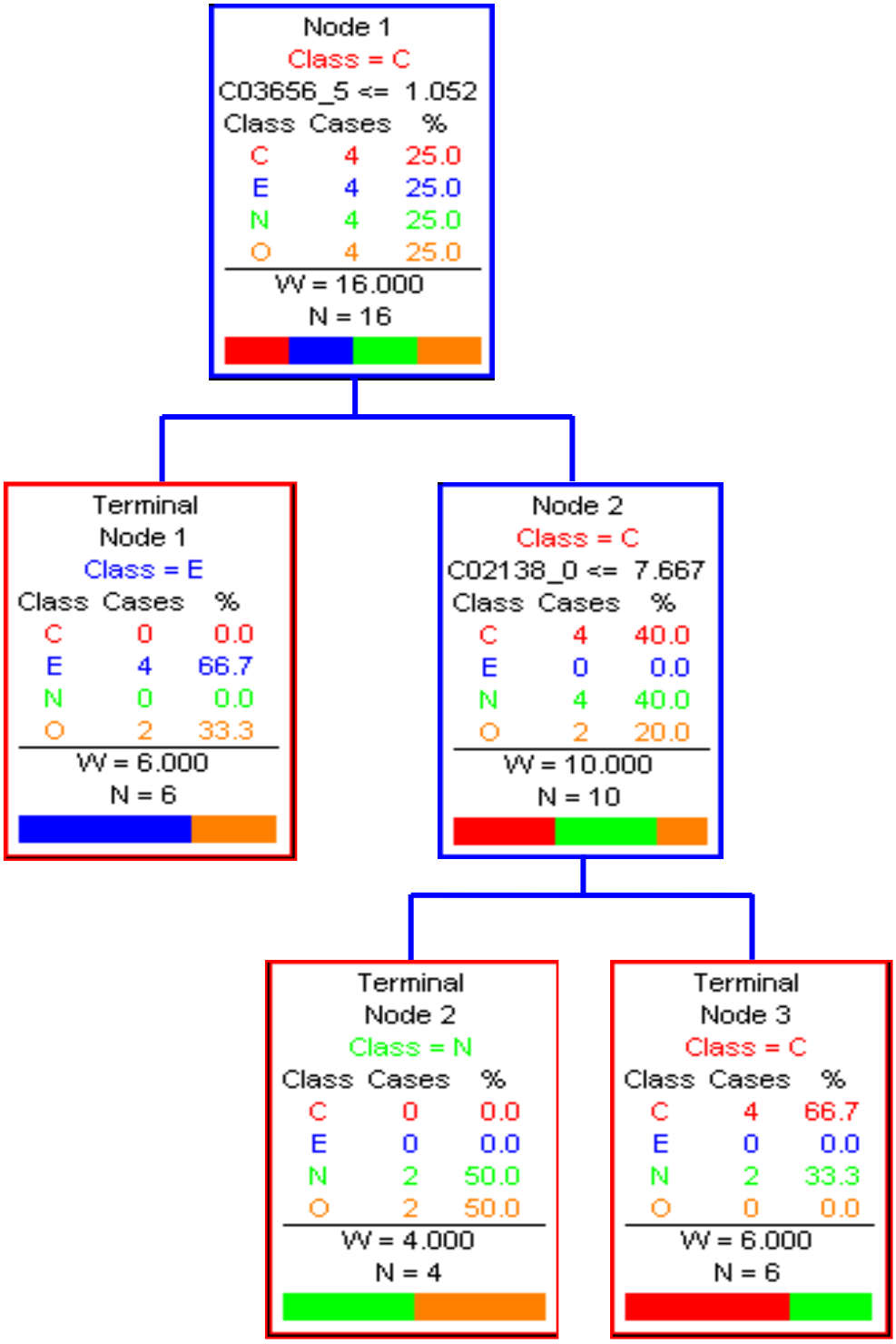
Two peaks, *m/z* 3656 and 2138 from CGB lectin captured urinary glycopeptides profile were chosen to set up the decision tree by the Biomarker Pattern Software 5.0 for group discrimination. Figure 3.14 shows the tree structure and the sample distribution for these peaks in the CGB lectin captured urinary glycopeptide profile. Samples having *m/z* 3656 with peak intensity lower than 1.052 μA went to Terminal node 1. At this node, four samples of ECa and two samples of OCa were selected. The remaining samples entered Node 2, which comprised four samples each of control and CCa, and two samples for OCa. At node 2, samples were divided based on *m/z* 2138, where samples with peak intensity higher than 7.667 μA went to Terminal node 3. Four samples of CCa and two samples of control fell into this terminal node. The remaining samples entered Terminal node 2, which were two samples each of control and OCa.

### 3.10.4 Statistical analysis of protein peaks that discriminate the groups

The significance of the peaks that discriminated each of the groups based on the decision tree constructed from the CMB lectin captured urinary glycopeptides profiles

**Figure 3.14: Tree structure profiles and samples distribution of CGB lectin captured urinary glycopeptides**

Peaks with *m/z* of 2138 and 3656 were chosen to set up the decision tree. C, E, N and O refer to CCa, ECa, control and OCa, accordingly. The root node and the terminal node were indicated as blue and red colour, respectively.


Node 1
Class = C
C03656_5 <= 1.052
Class Cases %
C 4 25.0
E 4 25.0
N 4 25.0
O 4 25.0
W = 16.000
N = 16
Terminal
Node 1
Class = E
Class Cases %
C 0 0.0
E 4 66.7
N 0 0.0
O 2 33.3
W = 6.000
N = 6
Node 2
Class = C
C02138_0 <= 7.667
Class Cases %
C 4 40.0
E 0 0.0
N 4 40.0
O 2 20.0
W = 10.000
N = 10
Terminal
Node 2
Class = N
Class Cases %
C 0 0.0
E 0 0.0
N 2 50.0
O 2 50.0
W = 4.000
N = 4
Terminal
Node 3
Class = C
Class Cases %
C 4 66.7
E 0 0.0
N 2 33.3
O 0 0.0
W = 6.000
N = 6

was determined with EDM program. The peaks *m/z* 1201 and 1449 obtained from CMB lectin captured urinary glycopeptides were significantly different when compared between groups with the *p*-value of 0.028. Figure 3.15 shows the representative spectral view of *m/z* 1201 and 1449 for the CMB lectin urinary glycopeptides profiles for controls and the cancer groups. Both proteins were overexpressed in all groups of cancers relative to the controls. Indeed, the mean peak intensity of *m/z* 1201 and 1449 in OCa samples were 137.034±17.399 μA and 164.353±15.821 μA, respectively, which were higher than the other two cancer cohorts. Table 3.10 shows the mean peak intensity of *m/z* 1201 and 1449 captured by the CMB lectin chip.

Similar statistical analysis was conducted for the CGB lectin captured urinary glycopeptides for the peak *m/z* 2198 and 3656. Figure 3.16 shows the spectral view of the peaks. However, both peaks were not significantly different between each of the groups ($p>0.05$). The mean peak intensity of both peaks were higher in the controls compared to the other three cancer groups, with the value of 4.884±3.207 μA and 13.849±7.413 μA, respectively. Table 3.11 shows the mean peak intensity of *m/z* 2198 and 3656.

**Table 3.10: Mean peak intensity for *m/z* 1201 and 1449 captured by CMB lectin chip**

| ***m/z*** | **Mean peak intensity (μA)** | | | |
|---|---|---|---|---|
| | Control | ECa | CCa | OCa |
| **1201** | 24.601±12.102 | 31.018±4.263 | 36.284±16.459 | 137.034±17.399 |
| **1449** | 59.299±15.740 | 118.036±10.443 | 98.098±31.132 | 164.353±15.821 |

All values were expressed as mean±S.E.M

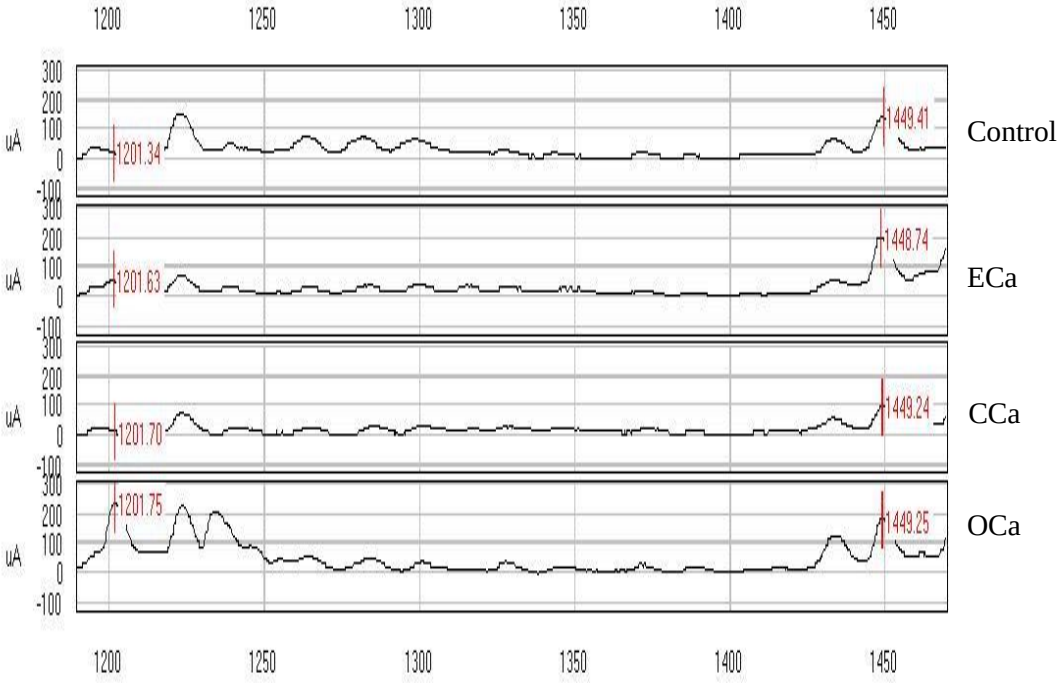



**Figure 3.15: Representative spectral view of *m/z* 1201 and 1449 for the CMB lectin urinary glycopeptides profiles for control and the cancer groups**

Peaks with *m/z* 1201 and 1449 were selected to construct the decision tree and thus subjected to the statistical analysis with Expression Difference Mapping. Both peaks were significantly different between each groups. The x axis represents *m/z* values and y axis represents relative intensity in μA.

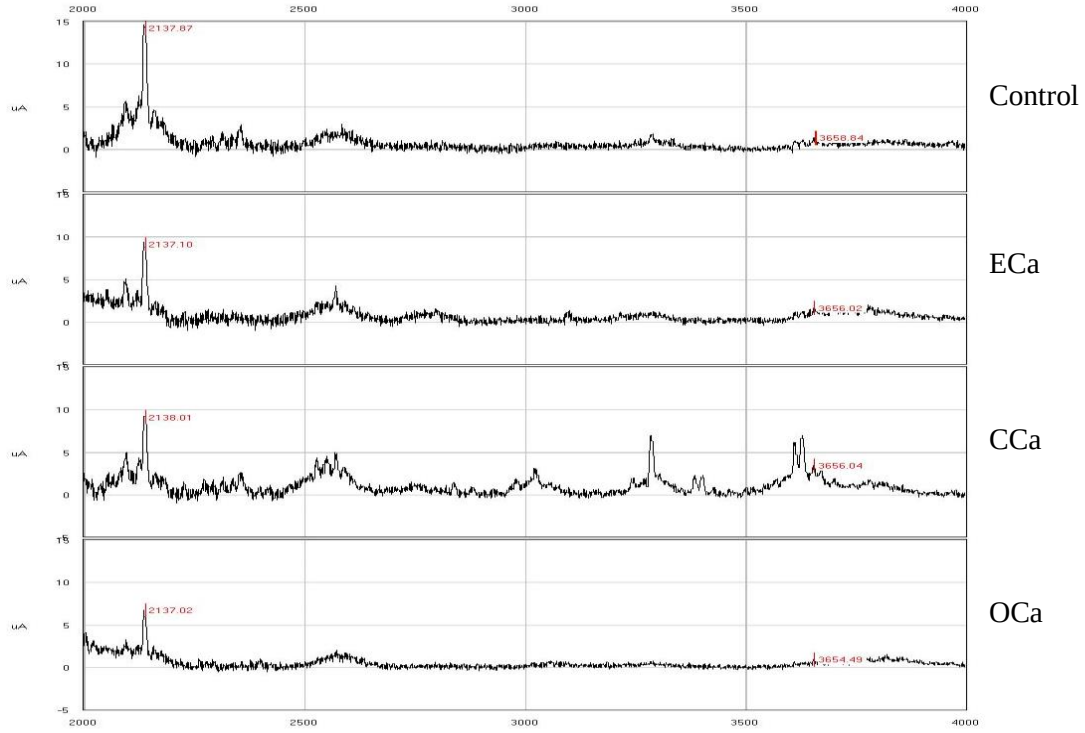



**Figure 3.16: Representative spectral view of *m/z* 2138 and 3656 for the CGB lectin urinary glycopeptides profiles for control and the cancer groups**

Peaks with *m/z* 2138 and 3656 were used to set up the decision tree even though they were not statistically significant when compared between the four groups. The x axis represents *m/z* values and y axis represents relative intensity in μA.

**Table 3.11: Mean peak intensity for *m/z* 2138 and 3656 captured by CGB lectin chip**

| *m/z* | Mean peak intensity (μA) | | | |
|---|---|---|---|---|
| | Control | ECa | CCa | OCa |
| **2138** | 13.849±7.413 | 7.597±2.653 | 12.628±2.795 | 5.529±3.065 |
| **3656** | 4.884±3.207 | 0.624±0.207 | 2.046±0.387 | 1.151±0.647 |

All values were expressed as mean±S.E.M

### 3.10.5 Sensitivity and specificity determination for the decision tree

A total of 16 samples with four samples from each group (controls and 3 cohort cancer patients) were used to evaluate the sensitivity and the specificity of the peak in discriminating the cohorts from each other. Based on the figure 3.15, two peaks with *m/z* of 1201 and 1449 were used to set up the decision tree for CMB lectin captured urinary glycopeptides. All ECa samples were correctly classified and 11 out of 12 non ECa samples were correctly classified. This result yielded a sensitivity of 100% and a specificity of 91.67%. On the other hand, the sensitivity and the specificity for OCa was 100% when all OCa samples were successfully classified and no samples for other cohorts were misclassified as OCa. Unlike ECa and OCa samples, decision tree was not able to classify the CCa since all CCa samples were grouped in the decision nodes. Table 3.12 shows the sensitivity and the specificity of each cohort in CMB lectin captured urinary glycopeptides. As shown in Figure 3.16, two peaks with *m/z* 2138 and 3656 were used to set up the decision tree for CGB lectin captured urinary glycopeptides. Based on the decision tree, all samples of ECa and CCa were classified as their own group, which yielded a sensitivity of 100% for each of them. However, their specificity were only 83.33% when two samples of non specified samples were misclassified into their group. Besides, the decision tree could not classify OCa

samples where all decision nodes and terminal nodes were not able to specify OCa samples. Table 3.13 shows the sensitivity and the specificity of each cohort in CGB lectin captured urinary glycopeptides.

**Table 3.12: Sensitivity and specificity of CMB lectin to discriminate urinary glycopeptides from patients with cancers using the SELDI-TOF-MS chip based assay**

| Sample | Cases | Number of correctly classified samples* | Classification accuracy (%) |
|---|---|---|---|
| ECa | 4 | 4 | 100[a] |
| Non-ECa | 12 | 11 | 91.67[b] |
| CCa | - | - | - |
| Non-CCa | - | - | - |
| OCa | 4 | 4 | 100[a] |
| Non-OCa | 12 | 12 | 100[b] |

*Determined from the decision tree of CMB lectin captured urinary glycopeptides
- Not applicable
[a] Sensitivity
[b] Specificity

**Table 3.13: Sensitivity and specificity of CGB lectin to dicriminate urinary glycopeptides from patients with cancers using the SELDI-TOF-MS chip based assay**

| Sample | Cases | Number of correctly classified samples* | Classification accuracy (%) |
|---|---|---|---|
| ECa | 4 | 4 | 100[a] |
| Non-ECa | 12 | 10 | 83.33[b] |
| CCa | 4 | 4 | 100[a] |
| Non-CCa | 12 | 10 | 83.33[b] |
| OCa | - | - | - |
| Non-OCa | - | - | - |

* Determined from the decision tree of CGB lectin captured urinary glycopeptides
- Not applicable
[a] Sensitivity
[b] Specificity

# Chapter 4
# Discussion

# CHAPTER 4: DISCUSSION

## 4.1 Collection and processing of urine samples

The procedures for collection and processing of the urine samples are important before a proteomic analysis because they may affect the results. Hence, several precautions were taken in the collection and processing steps. The morning midstream urine was chosen to be collected for the analyses instead of the first voided urine because it provides consistent protein content for profiling and characterization (Casado-Vela *et al.,* 2011). The addition of 20 mM sodium azide into the urine samples was done immediately after collection in order to prevent any bacterial and/or fungi growth in the urine samples. Presence of the microorganisms can interfere with the subsequent proteome analysis of the urine samples. Protease inhibitors were not added into the urine samples because proteases are deactivated once they are voided from the body (Coon *et al.,* 2008). Furthermore, the protease inhibitor might interfere with the subsequent proteome analysis. Before further processing, urine samples were centrifuged at 10000 × *g* for 15 minutes to eliminate contaminants that may be present in the urine, such as cells, microorganisms and debris. Removal of the contaminants could eventually reduce the appearance of additional substances in the urine samples, which sometimes affects the quality of the results of the proteome analysis. Subsequent to this, the supernatants were dialyzed against four changes of distilled water, which reduced the salt concentration of the urine samples, which otherwise would interfere with the 2-DE separation of the urinary proteins.

The control of temperature during urine sample processing and storage would significantly influenced the results achieved for the proteome analysis. Thus, all urine sample processing procedures were done within not more than three days of storage at

4°C, where protein degradation was still minimal. This was in accordance to the report by Schaub et al. (2004), which showed that urinary proteins were not significantly altered when stored up to 3 days at 4°C. After dialysis, the urine samples were lyophilized to prevent degradation of the urinary protein and growth of bacteria. Prior to protein analysis, the lyophilized urinary samples were stored at -80$^{o}$C to ensure that the intact structure of urinary proteins was maintained for long term storage. The dialysis and lyophilization of the urine sample could improve the reproducibility in term of variability in protein concentration caused by hydration status of each samples and give better resolution for its downstream analysis.

The concentrations of the urinary proteins obtained from patients were estimated using the standard curve constructed from a range of BSA dilutions. Generally, higher concentrations of urinary proteins were observed in cancer patients compared to the control subjects. The excretion of abnormal amount of proteins in the urine, which is referred as proteinuria, more often occurs in a number of pathological conditions, including cancer (Rudman *et al.,* 1978). This is to be expected as elevated amounts of protein are circulated in the body and finally deposited in the kidney and excreted as urine during the malignancy condition (Pedersen and Milman, 1996). In this study, small number of samples (control=11, ECa=7, OCa=9, CCa=8) was collected and used for the present gel- and lectin-based proteomic analyses because these analyses were tedious and expensive. Furthermore, it was difficult to obtain sufficient number of samples for the study.

## 4.2 2-DE analysis of urinary proteins obtained from controls and patients with ECa

Urine is one of the body biofluids which is commonly used in clinical diagnosis. Unlike serum or plasma, the collection of human urine is more convenient and non-invasive, which makes it one of the favourite choices for diagnostic purposes. Many studies have been done to explore the diagnostic value of urine for detecting cancer (Tantipaiboonwong *et al.,* 2005; Kreunin *et al.,* 2007). The search for cancer biomarkers from urine has gained momentum with the development of advanced proteomic technology. This include the application of 2-DE and mass spectrometry. 2-DE is a conventional method that is widely used to analyze a large number of proteins present in a given sample. The usage of this method enables the quantification of proteins and thus provides a platform to conduct a comparative study of protein expression between controls and patients. However, 2-DE has its own drawbacks such as difficulty to detect extremely low or high molecular weight proteins, membrane proteins and low abundance proteins. The combination of 2-DE with other proteomic techniques such as LC-MS/MS and SELDI-TOF could enhance the sensitivity in analyzing a complete proteome of a given sample because different proteomic technologies would generate different findings. Furthermore, their outcomes could also improve our understanding on the facilitation of the cancer management and therapeutic intervention.

In the present study, highly resolved 2-DE silver-stained urinary protein profiles were developed for controls and patients with ECa. Because of the inter-variation and heterogeneity of protein composition for each individual, the image analysis of the protein clusters were restricted to those proteins that appeared consistently in all

individuals. Therefore, seven protein clusters namely KNG, AAG, ZAG, CD59, AMBP, IG3C and IGKC were chosen for densitometry analysis. Mass spectrometry analysis with MALDI-TOF/TOF was performed to identify these proteins. Analysis of the 2-DE silver-stained profiles of controls and patients with ECa demonstrated the altered expression of AAG, ZAG and CD59. The intensity of AAG and ZAG were significantly higher and the expression of CD59 was much lower in patients with ECa. When the present data was compared to those previously reported on patients with OCa and CCa, it appears that ECa may be differentiated from OCa and CCa (Abdullah-Soheimi, 2010; Mu *et al.*,2012) based on the different protein expression. Each of these cancers seems to have their own unique urinary protein profile.

The altered expression of the urinary proteins for ECa detected in the present study were in slight contradiction to the previous study conducted by Abdul-Rahman and her colleagues on unfractionated serum samples (Abdul-Rahman *et al.,* 2007). In their analysis of serum samples, Abdul-Rahman el al. (2007) reported the differential expression of five high abundance proteins found in patients with ECa compared to controls. In their study, the expression of serum ZAG was not significantly different between the ECa patients and the controls, whilst the data of the present study showed higher expression of urinary ZAG in the ECa patients. Urine is the filtration of plasma/serum by the glomerulus, which is a selective process based on dimension, charge and configuration of the protein. Therefore, the divergence of data was probably due to the different types of samples used for both studies.

## 4.3 Purification of CMB and CGB lectins

Prior to the analysis of urinary glycoproteins, CMB and CGB lectins were isolated from the crude extracts of seeds obtained from the champedak fruit (*Artocarpus integer*). Before the isolation of the lectins, the champedak seeds were treated with several procedures in order to obtain optimum amount of lectins. Firstly, the seed coats were peeled off, prior to grinding. The reason for removing the seed coats was to remove impurities that might be present. Then, the seeds were homogenized to break the cell walls. This was followed by the centrifugation of the crude extracts, which eliminated the cellular debris. The crude lectins were precipitated with ammonium sulphate at 60% saturation. The hydrophilic property of the proteins was disrupted when sufficient ionic strength was exerted by an appropriate concentration of ammonium sulphate to the proteins.

The mannose and galactose sugar affinity chromatography columns were used to purify the CMB and CGB lectins, respectively, from the crude lectin extracts (Hashim *et al.,* 1991; Lim *et al.,* 1997). The unbound fractions from galactose sugar affinity chromatography columns were subjected to mannose sugar affinity chromatography column to isolate the CMB lectin. The lectins were eluted out when 0.8 M of mannose or galactose were applied into the mannose or galactose sugar affinity chromatography columns, respectively. This is due to the higher attraction of the lectins to the high concentration of sugars compared to the immobilized sugars.

The purity of the lectins was assessed in order to avoid presence of other proteins, which might affect the subsequent experiments. Both lectins were reduced and subjected to 18% SDS-PAGE for separation. Under the reducing condition, the

CMB lectin was resolved into a distinctive band, with molecular weight of approximately 14 kDa, while the CGB lectin was resolved into two bands, with approximate molecular weights of 14 kDa and 16 kDa. The appearances of the bands for the respective lectins were comparable with their proposed subunit (Lim *et al.,* 1997; Abdul Rahman *et al.,* 2002). The purity of the lectins were confirmed where there was no additional band detected in the SDS-PAGE gel.

## 4.4 2-DE lectin blot analysis of urinary proteins obtained from controls and patients with ECa, OCa and CCa

The present study not only examined the altered expression of the urinary proteins but was also extended to the study of their glycosylation. Glycosylation is one of the posttranslational modifications that occur when an oligosaccharide binds to the suitable binding site of a protein. It greatly alters the physical properties of proteins and their biological functions such as stabilizing the protein orientation, maintaining homeostasis of the cellular processes and acting as functional molecules. Many evidences indicate that abnormally glycosylated proteins are implicated with cancer progression (Abdul-Rahman *et al.,* 2007; Dai *et al.,* 2009; Dennis *et al.,* 1999; Fujimura *et al.,* 2008; Mu *et al.,* 2012). Thus, the study of the altered glycosylation could provide a meaningful information of its roles in cancer development.

The current study, which involved patients with ECa, CCa and OCa, unraveled the glycosylation of urinary proteins using the CGB and CMB lectins as probes. The lectin approach allowed the validation of the previous 2-DE silver-stained data and the detection of additional proteins that were either not detected or not well resolved in the 2-DE silver-stained method (Abdul-Rahman *et al.,* 2007). In addition, it was also

possible to determine whether urinary proteins were aberrantly glycosylated by comparing the 2-DE lectin blots with the similar 2-DE silver-stained urinary profiles of controls and patients.

To date, *N*- and *O*-glycosylated urinary protein maps have not been established, even though a number of complete urinary protein maps have been reported since year 1979 (Anderson *et al.,* 1979; Candiano *et al.,* 2010). Through the present study, two different comprehensive *O*- and *N*-glycosylated urinary protein profiles have been successfully developed using the enzyme conjugated CGB and CMB lectins, respectively. When the 2-DE separated proteins were transferred onto nitrocellulose membranes and probed with enzyme-conjugated CGB lectin, profiles consisting of only *O*-glycosylated urinary proteins were obtained. Comparative densitometry analyses were made between 2-DE *O*-glycosylated protein profiles generated from the urine samples of the controls and patients with ECa, OCa and CCa. The significant altered levels of a single *O*-glycoprotein, which were subsequently identified as a 51 kDa fragment of NEB via on-membrane digestion followed by mass spectrometry and database searching, were demonstrated in the *O*-glycosylated urinary protein profiles of patients with ECa. The expression of urinary NEB was not significantly different between patients with CCa or OCa compared to the controls.

However, in patients with CCa, none of the *O*-glycosylated urinary proteins that were detected was significantly different from those of the controls. In case of the OCa patients, reduced levels of KNG were detected compared to the controls. And this provides further confirmation of an earlier work performed by silver staining of 2-DE gels (Abdullah-Soheimi *et al.,* 2010). In addition, LRG and CLU were also found to be

differentially expressed in patients with OCa, although this was not detected earlier (Abdullah-Soheimi *et al.*, 2010). NEB, LRG and CLU may not have been detected or well resolved in the earlier 2-DE experiments probably because of their relatively low amount in the urine samples.

A profile consisting of only *N*-glycosylated urinary proteins was obtained when 2-DE separated proteins were probed with enzyme conjugated CMB lectin. Comparison of *N*- and *O*-glycosylated urinary proteins profiles showed that HPX, IGHA1, KNG, LRG, CLU, ITIH4, AMBP, TAGLN and HSPG2 were present in both profiles. These proteins contain both consensus sequence of Asn-X-Ser(Thr) and serine or threonine residues for the binding of *N*-acetylglucosamine and *N*-acetylgalactosamine, respectively. When image analysis of the 2-DE separated *N*-glycosylated urinary protein profiles was made between the controls and patients with ECa, the spots of ZAG and AAG appeared more intense in the patient samples. These findings are comparable to the results that were established by subjecting the 2-DE gels to silver-staining, which were discussed earlier.

Similarly, the intensities of ZAG and AAG were also much higher in the *N*-glycosylated urinary protein profile of patients with OCa but this was not earlier detected in the silver-stained gel analysis (Abdullah-Soheimi *et al.*, 2010). The noticeable difference in the intensities of both proteins between the *N*-glycosylated protein profiles and those generated by silver-stainining from the patients with OCa indicated that these proteins may be differentially glycosylated. This may be attributed to the increase or higher exposure of the mannosyl residues of the *N*-glycans of the

urinary ZAG and AAG, which led to the strong interaction of the CMB lectin to the glycans side chains of the proteins, although this requires further validation work.

Similar to those observed in the *O*-glycosylation profiles of patients with OCa, the altered expression of KNG, CLU and LRG in the *N*-glycosylated urinary protein profiles of patients with OCa was demonstrated in the present study. This implied their altered levels were most probably caused by their differential expression in the cancer state, although their fold changes were different in both profiles. In the case of HP, its differential expression was detected only with the CMB lectin in the *N*-glycosylation profiles of the OCa patients. When the protein spot was subjected to MS/MS analysis, its peptide sequence was identified as β chain of HP with high confidence. The current findings of the altered levels of LRG, CLU and HP β chain was compatible with the previous study on the sera of OCa (Chen *et al.*, 2008). It can be postulated that the increase of proteins in the urine of OCa patients may reflect the change that occurred in the serum. Similar to the *O*-linked glycosylation profile, no significant changes were observed in the *N*-glycosylation profiles of patients with CCa as compared to that of the controls. This may due to the large discrepancy of protein levels in the urine samples of CCa patients.

## 4.5 LC-MS/MS analysis of bound fractions of urinary glycoproteins obtained from CGB and CMB lectin affinity chromatography

Low abundance proteins could deliver a vast amount of clinical information regarding the pathogenesis of cancer. Owing to the enormous complexity of the urinary proteome, they are normally masked by the high abundance proteins, making them difficult to be detected. To a certain extent, lectin-affinity chromatography allows the

detection of low abundance proteins by reducing the complexity of the protein mixture where only glycosylated proteins was isolated. Glycoprotein enrichment was used because up to 50% of the human proteins are glycosylated (Budai *et al.,* 2009). However, unlike lectin-blotting, lectin-affinity chromatography captures glycoproteins in their native form, in which the glycoprotein functional structures are preserved. Therefore, this method would selectively bind to proteins with exposed glycans.

The analysis of samples using LC-MS/MS is highly sensitive and applicable for high throughput assessment although it is rather costly. At present, several reports have addressed the study of the *N*-glycosylated proteins in the urine using the lectin enrichments methods and followed by LC-MS/MS analysis. Yang and colleagues developed a glycoproteomic strategy for bladder cancer biomarker discovery using dual-lectin affinity enrichment of urinary *N*-glycosylated proteins followed by LC-MS/MS (Yang *et al.,* 2011). They discovered the elevation of alpha-1-antitrypsin in the urine of bladder cancer patients after quantifying urinary proteins using label free spectral counting method and validating it with ELISA. This suggested that alpha-1-antitrypsin could be a potential complementary biomarker for bladder cancer. At the same time, similar glycoproteomic approach established by Vivekanandan-Giri et al. (2011) was able to identify a few differentially expressed *N*-glycosylated proteins in the urine of patients with chronic kidney disease (Vivekanandan-Giri *et al.,* 2011). Wang et al. (2006) profiled urinary *N*-glycosylated proteins of healthy controls by using Con A lectin affinity chromatography with LC-MS/MS analysis. To the best of our knowledge, the current study is the first to focus on *O*-glycosylated urinary proteins by combining lectin enrichments and mass spectrometry analysis.

In this study, the CGB and CMB lectins were used for protein enrichment. LC-MS/MS analysis of CGB and CMB lectins bound fractions gave rise to the identification and characterization of 36 and 46 urinary proteins with high confidence, respectively. Almost all of the proteins detected previously from 2-DE silver-stained and lectin blot were detected in the bound fractions except CLU. Apart from identifying the proteins, the subcellular localizations of these proteins were also determined with the SWISS-Prot annotations. Most of the identified proteins originated from secreted proteins. Therefore, it is possible that these proteins may not be part of urine, but rather secreted from the host organ or tumor itself and excreted together with the urine. Meanwhile, a number of membrane proteins have also been resolved from both the lectin enrichments of bound fractions coupled with LC-MS/MS analysis as they were not able to be detected in previous 2-DE approach because of their low solubility. In terms of glycosylation, some proteins were not annotated to be glycosylated in the Uniprot database. Nevertheless, they were predicted as potentially glycosylated when their sequences were analyzed with NetOGlyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc/) and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). Both servers can determine whether the consensus sequences may be recognized by the glycosyltransferase as the acceptor site for glycans (Julenius *et al.*, 2005).

In the recent years, label free relative quantitation by spectral counting has been demonstrated to correlate with protein abundances (Liu *et al.*, 2004). Since all altered proteins detected previously were secreted proteins, it could be assumed that their levels should be equal to their glycosylated forms (Chen *et al.*, 2011). Hence, their total number of identified MS/MS spectra were counted and compared. However, an issue that arised from the spectra counting approach was that small proteins tend to have

fewer peptides identified per protein compared to large proteins. Thus, the spectra of those particular proteins were normalized with normalized spectral abundance factor (NSAF) in order to prevent false quantitative result caused by the discrepancy of the proteins' length or sequence (Florens *et al.,* 2006; Zybailov *et al.,* 2006). Correspondingly, the spectral counting data of LRG from CGB lectin and KNG from CMB lectin bound fractions validated their differential levels observed in the OCa patients' urine. For HP, seven out of 11 spectra detected in the CMB lectin chromatography bound fractions of OCa sample belonged to the β-chain, which was comparable to the differential expression that was observed in the CMB lectin blot analysis. Apart from that, the concentrations of ZAG and AAG detected in the bound fractions of the CMB lectin column from OCa patients were found to be unchanged. This conceived the possibility of aberrant *N*-glycosylation of both the proteins. There was a notable difference in Western blotting-CMB lectin analysis with the current result for these proteins. Nonetheless, this hypothesis was not conclusive as further study on their *N*-glycan structures are required to prove it.

In the CMB lectin bound fractions of patients with ECa, both the patient:control expression ratios of urinary ZAG and AAG were more than 1.9 fold. This is compatible with the altered expression of both the proteins that was detected in the previous 2-DE experiments using silver staining (Section 3.5.1). The aberrant expression of CD59 in the three cohorts of cancer was affirmed by detection of two corresponding spectra exclusively in the CMB bound fractions of the controls but not in the patients (Abdullah-Soheimi *et al.,* 2010; Abdullah-Soheimi, 2012; Mu *et al.,* 2012). Similar analysis on the CGB bound fractions of ECa further confirmed the low level of the NEB fragment in the urine of the patients. The deviation protein expression between individual subjects lead to the disparity of the magnitude of fold changes for the altered

proteins between the spectral counting method and 2-DE profiles. Therefore, there is a need to further validate the present biomarker candidates using large scale investigations using a simpler technique such as ELISA for the clinical purpose of early detection of ECa, OCa and CCa.

## 4.6 Biological functions of urinary proteins that were differentially expressed

In this section, the urinary glycoproteins whose levels were altered in patients with ECa, OCa and CCa that were detected in the present study is discussed with reference to their structures, biological functions and possible reasons for their abnormal expression. Apart from that, the possibility of the urinary proteins being aberrantly glycosylated is also discussed.

### 4.6.1 Zinc alpha-2 glycoprotein (ZAG)

ZAG is a glycoprotein with molecular weight of approximately 40 kDa, secreted by a variety of epithelial cells (Hale *et al.*, 2001). It is generally described as a stimulator for lypolysis although its function may include as a carrier protein, immunoregulator and cell adhesion molecule (Hassan *et al.*, 2008). ZAG may also play a role in cancer where its altered expression in the sera was widely documented in patients with prostate, breast and cervical cancers (Hale *et al.*, 2001; Abdul-Rahman *et al.*, 2007; Dubois *et al.*, 2010). The protein may be associated with pathways involving adipokines and estrogen, and its differential expression in patients with ECa was postulated to be due to the changes in the levels of these factors (Dubois *et al.*, 2010). In addition, ZAG has also been suggested to play a role in the development of cancer cachexia, one of the stimulator of carcinogenesis (Gordon *et al.*, 2005).

### 4.6.2 Alpha-1 acid glycoprotein (AAG)

AAG, also known as orosomucoid, is a heterogeneous protein with a variety of important roles in binding and transportation of endogenous basic substances and drugs, and in an immunomodulatory function (Budai *et al.*, 2009). It is also a positive acute phase protein as its concentration can elevate several folds during inflammation. Its up-regulation in serum was documented in several types of cancer including ECa and OCa (Tosner *et al.*, 1988; Wang *et al.*, 2011). It is a highly *N*-glycosylated, with the glycans constituting about 45% of its molecular mass of 41 kDa (Herve *et al.*, 1998). Since the intensity pattern of the protein spots detected in the 2-DE gel and CMB lectin blot was comparable in patients with ECa, it could be concluded that the glycosylation of the protein was most probably normal. On the other hand, the possible changes in *N*-glycosylation of AAG was suggested in the patients with OCa in the present study when the expression of the protein appeared to be altered in the CMB lectin blot but not in silver stained 2-DE gels (Abdullah-Soheimi *et al.*, 2010). And this is consistent with the findings from Saldova and colleagues, which indicated the changes in the AAG glycoforms, which contributed to changes of the serum glycome of patients with OCa (Saldova *et al.*, 2007).

### 4.6.3 CD59

CD59, also known as protectin, is a cell surface protein that functions to inhibit the membrane attack complex of the complement pathway. The present study has demonstrated the decreased in excretion of CD59 in the urine of patients with the three groups of cancer patients studied. A similar excretion profile of CD59 in the urine had been previously reported in pancreatic ductal adenocarcinoma and bladder cancer (Kreunin *et al.*, 2007; Weeks *et al.*, 2008). Its altered excretion in the serum/plasma has

never been reported even though the altered expression of CD59 in malignant tissues and in the urine had been observed (Murray *et al.*, 2000). This might be due to its low solubility and thus difficult to be detected in the serum and plasma. The reason behind the reduced expression of CD59 in gynaecological cancers is not known, but the depletion of CD59 by cancer cells to resist the attack of complement might explain why less of this protein are being solubilized and excreted in the urine (Abdullah-Soheimi *et al.*, 2010).

### 4.6.4 Nebulin (NEB)

Nebulin, a protein with a deduced molecular weight of 772.9 kDa, functions as a template for polymerization of actin (Pappas *et al.*, 2011). The protein is expressed predominantly in the thin filaments of striated muscle. It is known to be glycosylated, although the precise structure of its glycan moiety has never been characterized (Madera *et al.*, 2006). Hence, the result of the CGB lectin analyses performed in this study is a form of evidence that nebulin is *O*-glycosylated. The low molecular weight of the nebulin spot detected in the 2-DE experiments is indicative of a truncated or cleaved protein. The reduced levels of tissue nebulin are commonly associated with myopathy (Ottenheijm *et al.*, 2010). However, its increased levels in the pancreatic juice are believed to be associated with the cellular turnover in patients with pancreatic cancer (Park *et al.*, 2011). To the best of our knowledge, there is no previous report on the detection of nebulin or its fragment in the urine, but the protein has been detected in other biofluids such as serum and pancreatic juice (Tanaka *et al.*, 2006; Park *et al.*, 2011).

### 4.6.5 Leucine rich alpha-2 glycoprotein (LRG)

LRG is a secreted protein with five glycosylation sites, which contains an *O*-linked oligosaccharide and four *N*-linked oligosaccharides chains (Rodriguez-Pineiro *et al.*, 2004). It is composed of leucine rich repeats, in which 17% of its total amino acid composition are leucine residues. Its actual functions are still unknown, although numerous studies have reported its roles in cell adhesion, granulocytic differentiation, cell migration, cell survival, protein interaction and signal transduction (Andersen *et al.*, 2010). In addition, it was also proposed as an acute phase protein when its level was reported to be elevated in microbial infections and certain cancers including OCa (Andersen *et al.*, 2010). LRG is mainly produced by the liver cell in response to OCa, but its production by the OCa cells alone had also documented (Boylan *et al.*, 2010). Thus, the secretion of LRG by the liver cells and OCa cells may contribute to the increase of LRG level excreted in urine from OCa patients.

### 4.6.6 Clusterin (CLU)

CLU, also known as apolipoprotein J, is a highly conserved heterodimeric secreted glycoprotein. It comprises 449 amino acids with two chains (α and β) of approximately 40 kDa each, linked by disulfide bonds. The protein has been implicated in various physiological processes, including lipid transportation, immune system regulation, membrane recycling, senescence, DNA repair, cell aggregation and cell adhesion (Rosenberg and Silkensen 1995; Trougakos and Gonos 2002). Moreover, the expression of CLU has also been reported to correlate with cell stress response, tumorigenesis and apoptosis, particularly in protecting cancer cells from complement-mediated lysis (Shannan *et al.*, 2006; Nafee *et al.*, 2012). The overexpression of CLU has been described in several types of cancers including OCa but the decrease of its

expression had been observed in prostate cancer and esophageal squamous cell carcinoma (Redondo *et al.,* 2000; Zhang *et al.,* 2003; Doustjalali *et al.,* 2004; Albert *et al.,* 2007; Chen *et al.,* 2008; Rizzi and Bettuzzi 2009). The present data derived from the profiles of glycoproteins detected using CGB and CMB lectins have demonstrated the increased levels of CLU in the urine of OCa patients and its expression was further statistically confirmed with competitive ELISA as well as Western blotting using monoclonal anti-CLU. In addition, the present study demonstrates for the first time that CLU may contain *O*-glycans since the protein was shown to interact with the CGB lectin. Based on its known structure, CLU may contain one *O*-glycan moiety at its potential *O*-glycosylation site at the amino acid position of 403 (Kapron *et al.,* 1997). On the other hand, the CLU from OCa patients may also be possibly differently *O*-glycosylated.

### 4.6.7 Kininogen (KNG)

KNG is a protein containing two isoforms produced by alternative mRNA splicing (Takagaki *et al.,* 1985). The KNG that was observed in the present study was actually the light chain component of cleaved high molecular weight kininogen. Numerous studies of KNG have demonstrated that it was significantly reduced in patients with gastrointestinal cancer, breast cancer and cervical cancer (Roeise *et al.,* 1990; Doustjalali *et al.,* 2004; Abdul-Rahman *et al.,* 2007). The present data further confirmed its expression level in the urine of the OCa patients, as reported by Abdullah-Soheimi et al. (2010). Reduced levels of KNG in the urine of OCa patients may reflect its anti-angiogenic property and inhibitory action on the proliferation of endothelial cells (Liu *et al.,* 2008).

### 4.6.8 Haptoglobin (HP)

HP is an acute phase protein existing as a disulfide-linked α2β2 heterotetramer. Polymorphism of HP arises from the combination of different α-chains with identical β -chains (Zhao *et al.,* 2007). These two chains are formed from the cleavage of a common precursor polypeptide during post-translational modification (Haugen *et al.,* 1981). The main function of HP is to bind free hemoglobin for protection from oxidative stress and to facilitate the hemoglobin uptake by the hemoglobin scavenger receptor CD163 (Zhao *et al.,* 2007; Kang *et al.,* 2011). Besides, it also plays important roles in angiogenesis, the regulation of epidermal cell transformation, the inhibition of prostaglandin synthesis and reverses cholesterol transportation (Abdullah *et al.,* 2009). The production of HP is mainly by the liver in large amount in response to the stimulation of infections, inflammations and various malignant diseases, including OCa, lung cancer, breast cancer, pancreatic cancer, esophageal squamous cell carcinaoma and bladder cancer (Zhao *et al.,* 2007). The present study has demonstrated the differential expression of HP-β chain from patients with OCa detected with CMB lectin, and this is in agreement with the study done by Ahmed et al (2004). However, the failure of HP-α chain to be detected with the same lectin was expected since none of the consensus sequences of *N*-glycosylation site is located in the α-chain (Fujimura *et al.,* 2008). Although the elevation of HP-β chain in OCa was most probably related to the acute phase response, it is plausible that the elevation was also due to its action as a non-specific immune suppressor of cancer (Abdullah *et al.,* 2009). Furthermore, as in the case of AAG that was mentioned earlier (Section 4.6.2), the increase of HP-β chain level could also account for the changes in the serum glycome such as the $SLe^x$ structure (Saldova *et al.,* 2007).

## 4.7 SELDI-TOF analysis of urinary glycopeptides of patients with ECa, OCa and CCa

SELDI-TOF has emerged as one of the most powerful tool for detection of potential biomarkers for diseases. In the technique, the SELDI-TOF protein chip selectively captures specific proteins, which thereafter were resolved by mass spectrometry. The application of SELDI-TOF has numerous advantages including its ease of use, high-throughput application and its capability to analyze small amounts of multiple samples in a short time. However, SELDI-TOF is not sensitive towards high molecular weight proteins and thus confined to the analysis of low molecular weight peptides. Furthermore, it is also unable to precisely identify proteins from the peak detected.

The serum of patients with ECa, OCa and CCa has been widely explored using the SELDI-TOF technology. From the serum spectra generated on the weak cation exchange (WCX2) protein chips, several researchers have previously identified potential biomarkers for ECa, OCa and CCa (Zhang *et al.,* 2006; Xia *et al.,* 2008; Zhu *et al.,* 2008). Nevertheless, urine samples have never been explored using the SELDI-TOF for these cancers although studies on the urine of patients with bladder cancer have been previously reported (Chen *et al.,* 2010).

In the present pilot study, a method utilizing immobilized CGB and CMB lectins preactivated protein chip array was developed to selectively capture *O*- and *N*-glycosylated urinary proteins for SELDI-TOF analysis. In view of the limitation of SELDI-TOF on the analysis of low molecular weight peptides, trypsin-digested urinary glycopeptides were analysed instead of crude urinary glycoproteins. Analysis of lectin-

captured glycopeptides generates profiles comprising subsets of the whole glycoproteome, which are selective to the type of glycans they contain. Therefore, the CGB and CMB lectins pre-activated protein chip assays that were developed in the present study were not only able to identify the urinary glycoproteins that may be differentially expressed, but also to determine the possibility of detecting proteins that may be aberrantly glycosylated.

In case of *O*-linked urinary glycopeptides captured by the CGB lectin coupled Protein Chip array, glycopeptides mass with *m/z* 2138 and 3656 were autoselected to form the decision tree by the Biomarker Pattern Software. However, the intensity of the peaks were not significantly different between the cancers, which was most probably due to the small number of samples used for the statistical analysis.

Similarly, when *N*-linked urinary glycopeptides were captured by the CMB lectin coupled Protein Chip array, glycopeptides mass with *m/z* 1201and 1449 were autoselected to form the decision tree by the Biomarker Pattern Software. The peak *m/z* 1201 apparently has a strong discriminatory power to distinguish OCa from CCa and ECa and its sensitivity and specificity were 100%. It is therefore a potential biomarker candidate for OCa. On the other hand, the peak *m/z* 1449 had the capability of discriminating ECa from the other two types of cancers. When peaks *m/z* 1201 and 1449 were subjected to statistical analysis using EDM, they were discerned to be significantly different between each type of cancer. Hence, in the present preliminary study, a combination of the two glycopeptides peaks that were detected using the CMB lectin coupled Protein Chip suggest the possibility of using the lectin-based SELDI-TOF assay to determine the three different types of gynaecological cancers although this

needs to be further validated using epidemiologically-represented sample sizes. In addition, the two glycopeptides peaks can be identified using available bioinformatic tools in the future.

# Chapter 5
# Conclusion

# CHAPTER 5: CONCLUSION

Subjecting urine samples of control subjects and patients with ECa to the analysis by 2-DE in the present study showed the significant differential expression of ZAG, AAG and CD59. When the analysis was repeated using CGB lectin to detect O-glycosylated proteins, NEB was further detected to be significantly reduced in the ECa patients compared to the controls. A similar analysis of the urinary *O*-glycosylated protein profiles of patients with OCa demonstrated the altered expression of KNG, CLU and LRG. In case of patients with CCa, however, the expression levels of all *O*-glycosylated proteins appeared to be comparable to those of the controls.

When the CMB lectin was used to generate a 2-DE profile consisting of exclusively *N*-glycosylated proteins, significant increase of ZAG and AAG was detected in the urine of ECa patients, which is similar to the results that was generated by silver staining of the 2-DE gels. In case of the OCa patients, comparative analysis of the *N*-glycosylated protein profiles showed overexpression of AAG, ZAG, LRG, CLU and HPβ chain but decreased in the levels of KNG. Conversely, the expression of *N*-glycosylated urinary proteins was comparable between the controls and CCa patients.

Analysis of bound fractions obtained from immobilized CGB lectin affinity chromatography by LC-MS/MS generated a list of *O*-glycosylated/potentially *O*-glycosylated proteins. The label free quantification analysis demonstrated the altered levels of LRG and NEB in patients with OCa and ECa, which is similar to the above-mentioned earlier findings by Western blot. Similarly, when immobilized CMB lectin affinity chromatography was used to isolate native *N*-glycosylated proteins, spectral counting showed altered levels of AAG and ZAG for patients with ECa and OCa and

differential expression of KNG and HPβ chain for OCa patients. These are also in agreement with the earlier findings by Western blot. In addition the LC-MS/MS analysis detected presence of CD59 peptides, which only appeared in the CMB lectin bound fractions from the control subjects, but not in the bound fractions collected from patients with ECa, OCa and CCa.

Analysis of the CMB lectin captured urinary glycopeptides of control subjects and patients with ECa, OCa and CCa by SELDI-TOF unraveled two peaks *m/z* 1201 and 1449 as group discriminators. Similar analysis using Biomarker Wizard performed on the CGB lectin captured urinary glycopeptides of control subjects and three groups of cancer patients determined *m/z* 2138 and 3656 as potential group discriminator.

The data of this study are suggestive of the potential use of urinary proteins such as ZAG, AAG, NEB, KNG, LRG, CLU, HPβ chain and CD59 as complementary biomarkers for ECa, OCa and CCa. Diagnostic kits using antisera against the proteins may be easily generated and used in the clinical setting. However, this requires further extensive validation on clinically representative populations. Such a study cannot be possibly performed using the present gel-based proteomic and lectin-based proteomic analyses approaches since they are laborious and expensive. Nevertheless, with the identification of the potential urinary biomarkers in the present study, validation is easily carried out in a large-scale investigation using assays like the protein array or multiplex ELISA.

# Reference

**REFERENCE**

Abdul Rahman, M., Anuar Karsani, S., Othman, I., Abdul Rahman, P. S. & Hashim, O. H. (2002). Galactose-binding lectin from the seeds of champedak (*Artocarpus integer*): sequences of its subunits and interactions with human serum O-glycosylated glycoproteins. *Biochem Biophys Res Commun* **295**(4), 1007-1013.

Abdullah, F. & Su, T. T. (2010). Enhancement of the cervical cancer screening program in Malaysia: a qualitative study. *Asian Pacific J Cancer Prev* **11**, 1359-1366.

Abdullah, M., Schultz, H., Kähler, D., Branscheid, D., Dalhoff, K., *et al.* (2009). Expression of the acute phase protein haptoglobin in human lung cancer and tumor-free lung tissues. *Pathol Res Pract* **205**(9), 639-647.

Abdullah-Soheimi, S. S., Lim, B. K. Hashim, O. H. & Shuib, A. S. (2010). Patients with ovarian carcinoma excrete different altered levels of urine CD59, kininogen-1 and fragments of inter-alpha-trypsin inhibitor heavy chain H4 and albumin. *Proteome Sci* **8**(1), 58.

Abdullah-Soheimi, S.S. (2012). Proteomic analysis of urinary proteins from patients with ovarian cancer and cervical cancer. Dissertation, University of Malaya, Kuala Lumpur.

Abdul-Rahman, P. S., Lim, B. K. & Hashim, O. H. (2007). Expression of high-abundance proteins in sera of patients with endometrial and cervical cancers: analysis using 2-DE with silver staining and lectin detection methods. *Electrophoresis* **28**(12), 1989-1996.

Agron, I. A., Avtonomov, D. M., Kononikhin, A. S., Popov, I. A., Moshkovskii, S. A., *et al.* (2010). Accurate mass tag retention time database for urine proteome analysis by chromatography-mass spectrometry. *Biochemistry* **75**(5), 636-641.

Ahmed, N., Barker, G., Oliva, K. T., Hoffmann, P., Riley, C., *et al.* (2004). Proteomic-based identification of haptoglobin-1 precursor as a novel circulating biomarker of ovarian cancer. *Br J Cancer* **91**(1), 129-140.

Albert, J. M., Gonzalez, A., Massion, P. P., Chen, H., Olson, S. J., *et al.* (2007). Cytoplasmic clusterin expression is associated with longer survival in patients with resected non small cell lung cancer. *Cancer Epidemiol Biomarkers Prev* **16**(9), 1845-1851.

Andersen, J. D., Boylan, K. L., Jemmerson, R., Geller, M. A., Misemer, B., *et al.* (2010). Leucine-rich alpha-2-glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients. *J Ovarian Res* **3**, 21.

Anderson, N. G., Anderson, N. L. & Tollaksen, S. L., *et al.* (1979). Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. *Clin Chem* **25**(7), 1199-1210.

Anderson, N. L. & Anderson, N. G. (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**(11), 1853-1861.

Anderson, N. L., Matheson, A. D. & Steiner, S. (2000). Proteomics: applications in basic and applied biology. *Curr Opin Biotechnol* **11**(4), 408-412.

Basu, P., Chandna, P., Bamezai, R. N., Siddigi, M., Saranath, D., *et al.* (2011). MassARRAY spectrometry is more sensitive than PreTect HPV-Proofer and consensus PCR for type-specific detection of high-risk oncogenic human papillomavirus genotypes in cervical cancer. *J Clin Microbiol* **49**(10), 3537-3544.

Beddy, P., O'Neill, A. C., Yamamoto, A. K., Addley, H. C., Reinhold, C., *et al.* (2012). FIGO Staging System for Endometrial Cancer: Added Benefits of MR Imaging. *Radiographics* **32**(1), 241-254.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Roy Statist Soc Ser B (Methodological)* **57**(1), 289-300.

Bewtra, C., Watson, P., Conway, T., Read-Hippee, C. & Lynch, H. T. (1992). Hereditary ovarian cancer: a clinicopathological study. *Int J Gynecol Pathol* **11**(3), 180-187.

Blackstock, W. P. & Weir, M. P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* **17**(3), 121-127.

Bonfils, E., Depierreux, C., Midoux, P., Thuong, N. T., Monsigny, M., *et al.* (1992). Drug targeting: synthesis and endocytosis of oligonucleotide-neoglycoprotein conjugates. *Nucleic Acids Res* **20**(17), 4621-4629.

Booth, M., Beral, V. & Smith, P. (1989). Risk factors for ovarian cancer: a case-control study. *Br J Cancer* **60**(4), 592-598.

Boulet, G. A., Horvath, C. A., Berghmans, S. & Bogers, J. (2008). Human papillomavirus in cervical cancer screening: important role as biomarker. *Cancer Epidemiol Biomarkers Prev* **17**(4), 810-817.

Boylan, K. L., Andersen, J. D., Andersen, L. B., Higgins, L. & Skubitz, A. P. (2010). Quantitative proteomic analysis by iTRAQ(R) for the identification of candidate biomarkers in ovarian cancer serum. *Proteome Sci* **8**, 31.

Brown, L. J. R. (2001). Histological Typing of Ovarian Tumours,: 2nd ed. Scully RE. *J Clin Pathol* **54**(7), 575.

Budai, L., Ozohanics, O., Ludányi, K., Drahos, L., Kremmer, T., *et al.* (2009). Investigation of genetic variants of alpha-1 acid glycoprotein by ultra-performance liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **393**(3), 991-998.

Cadieux, P. A., Beiko, D. T., Watterson, J. D., Burton, J. P., Howard, J. C., *et al.* (2004). Surface-enhanced laser desorption/ionization-time of flight-mass spectrometry (SELDI-TOF-MS): a new proteomic urinary test for patients with urolithiasis. *J Clin Lab Anal* **18**(3), 170-175.

Candiano, G., Santucci, L., Petretto, A., Bruschi, M., Dimuccio, V., *et al.* (2010). 2D-electrophoresis and the urine proteome map: where do we stand? *J Proteomics* **73**(5), 829-844.

Casado-Vela, J., Gómez del Pulgar, T., Cebrián, A., Alvarez-Ayerza, N. & Lacal, J. C. (2011). Human urine proteomics: building a list of human urine cancer biomarkers. *Expert Rev Proteomics* **8**(3), 347-360.

Chen, R., Tan, Y., Wang, M., Wang, F., Yao, Z., *et al.* (2011). Development of glycoprotein capture-based label-free method for the high-throughput screening of differential glycoproteins in hepatocellular carcinoma. *Mol Cell Proteomics* **10**(7), 7.

Chen, Y. T., Chen, C. L., Chen, H. W., Chung, T., Wu, C. C., *et al.* (2010). Discovery of Novel Bladder Cancer Biomarkers by Comparative Urine Proteomics Using iTRAQ Technology. *J Proteome Res* **9**(11), 5803-5815.

Chen, Y., Lim, B. K., Peh, S. C., Abdul-Rahman, P. S. & Hashim, O. H. (2008). Profiling of serum and tissue high abundance acute-phase proteins of patients with epithelial and germ line ovarian carcinoma. *Proteome Sci* **6**, 20.

Cole, H. B., Ezzell, J. W. Jr., Keller, K. F. & Doyle, R. J. (1984). Differentiation of *Bacillus anthracis* and other Bacillus species by lectins. *J Clin Microbiol* **19**(1), 48-53.

Conrads, T. P., Zhou, M., Petricoin, E. F., Liotta, L. & Veenstra, T. D. (2003). Cancer diagnosis using proteomic patterns. *Expert Rev Mol Diagn* **3**(4), 411-420.

Coon, J. J., Zurbig, P., Dakna, M., Dominiczak, A. F., Decramer, S., *et al.* (2008). CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin Appl* **2**(7-8), 964.

Creasman, W. (2009). Revised FIGO staging for carcinoma of the endometrium. *Int J Gynaecol Obstet* **105**(2), 3.

Dai, Z., Zhou, J., Qiu, S. J., Liu, Y. K. & Fan, J. (2009). Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. *Electrophoresis* **30**(17), 2957-2966.

Dall'olio, F. (1996). Protein glycosylation in cancer biology: an overview. *Clin Mol Pathol* **49**(3), M126-135.

Daly, M. & Obrams, G. I. (1998). Epidemiology and risk assessment for ovarian cancer. *Seminars in Oncology* **25**(3), 255-264.

Damme, E. M. (2008). Plant Lectins as Part of the Plant Defense System Against Insects. *Induced Plant Resistance to Herbivory*. A. Schaller, Springer Netherlands**,** 285-307.

Decramer, S., Gonzalez de Peredo, A., Breuil, B., Mischak, H., Monsarrat, B., *et al.* (2008). Urine in clinical proteomics. *Mol Cell Proteomics* **7**(10), 1850-1862.

Dennis, J. W., Granovsky, M. & Warren, C. E. (1999). Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **6**(1), 21-34.

Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M. L. & Kerbel, R. S. (1987). Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* **236**(4801), 582-585.

Diaz, C. L., Melchers, L. S., Hooykaas, P. J. J., Lugtenberg, B. J. J. & Kijne, J. W. (1989). Root lectin as a determinant of host-plant specificity in the Rhizobium-legume symbiosis. *Nature* **338**(6216), 579-581.

Doustjalali, S. R.,Yusof, R., Yip, C. H., Looi, L. M., Pillay, B., *et al.* (2004). Aberrant expression of acute-phase reactant proteins in sera and breast lesions of patients with malignant and benign breast tumors. *Electrophoresis* **25**(14), 2392-2401.

Doyle, R. J., Nedjat-Haiem, F., Keller, K. F. & Frasch, C. E. (1984). Diagnostic value of interactions between members of the family Neisseriaceae and lectins. *J Clin Microbiol* **19**(3), 383-387.

Dubois, V., Delort, L., Mishellany, F., Jarde, T., Billard, H., *et al.* (2010). Zinc-alpha2 glycoprotein: a new biomarker of breast cancer? *Anticancer Res* **30**(7), 2919-2925.

Duk, M. & Lisowska, E. (1981). *Vicia graminea* anti-*N* lectin: partial characterization of the purified lectin and its binding to erythrocytes. *Eur J Biochem* **118**, 131-136.

Durand, G. & Seta, N. (2000). Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin Chem* **46**(6 Pt 1), 795-805.

Engelmann, B. (1993). ABH antigens as recognition sites for the activation of red blood cell anion exchange by the lectin *Ulex europaeus* agglutinin I. *J Cell Physiol* **157**(2), 403-407.

Etzler, M. E. (1979). Lectins as probes in studies of intestinal glycoproteins and glycolipids. *Am J Clin Nutr* **32**(1), 133-138.

Farrell, F. & Lee, A. (2004). The erythropoietin receptor and its expression in tumor cells and other tissues. *Oncologist* **5**, 18-30.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., *et al.* (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**(12), 2893-2917.

Fernandes, B., Sagman, U. Auger, M., Demetrio, M. & Dennis, J. W. (1991). Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res* **51**(2), 718-723.

FIGO (International Federation of Gynecologist and Obstetricians). Cancer Committee. (1987) Changes in FIGO staging. *Obstetrics and Gynecology*, **138**.

Fliser, D., Novak, J., Thongboonkerd, V., Argilles, A., Jankowski, V., *et al.* (2007). Advances in urinary proteome analysis and biomarker discovery. *J Am Soc Nephrol* **18**(4), 1057-1071.

Florens, L., Carozza, M. J., Swanson, S. K., Fournier, M., Coleman, M. K., *et al.* (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**(4), 303-311.

Franco, E. L., Duarte-Franco, E. & Ferenczy, A. (2001). Cervical cancer: epidemiology, prevention and the role of human papillomavirus infection. *CMAJ* **164**(7), 1017-1025.

Fujimura, T., Shinohara, Y., Tissot, B., Pang, B. C., Kurogochi, M., *et al.* (2008). Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *Int J Cancer* **122**(1), 39-49.

Gabrielsen, M., Abdul-Rahman, P. S., Isaacs, N. W., Hashim, O. H. & Cogdell, R. J. (2010). Crystallization and initial X-ray diffraction analysis of a mannose-binding lectin from champedak. *Acta Crystallogr Sect F* **66**(5), 592-594.

Gabrielsen, M., Riboldi-Tunnicliffe, A., Abdul-Rahman, P. S., Mohamed, E., Ibrahim, W. I., *et al.* (2009). Crystallization and preliminary structural studies of champedak galactose-binding lectin. *Acta Crystallogr Sect F* **65**(9), 895-897.

Gadducci, A., Marrai, R., Baicchi, U., Gagetti, O., Facchini, V., *et al.* (1996). Prothrombin fragment F1+2 and thrombin-antithrombin III complex (TAT) plasma levels in patients with gynecological cancer. *Gynecol Oncol* **61**(2), 215-217.

Gallion, H. H., Pieretti, M., DePriest, P. D. & van Nagell, J. R. Jr. (1995). The molecular basis of ovarian cancer. *Cancer* **76**(10 Suppl), 1992-1997.

Gevensleben, H., Serce, N. & Buttner, R. (2010). Hereditary breast and ovarian cancers. *Pathologe* **31**(6), 438-444.

Ghazarian, H., Idoni, B. & Oppenheimer, S. B. (2011). A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem* **113**(3), 236-247.

Gluzman, D. F., Nadgornaia, V. A., Skliarenko, L. M., Abramenko, I. V., Sidorenko, S. P., *et al*. (1991). Immunoenzyme cytochemical methods, monoclonal antibodies and lectins in the diagnosis of tumors of the lymphoid tissue in children. *Pediatriia* **11,** 18-22.

Goldstein, I. J. & Poretz, R. D. (1986). The Lectins: Properties, Functions, and Applications in Biology and Medicine. *Academic Press* 33-247.

Goldstein, I. J., Hughes, R. C., Monsigny, M., Osawa, T. & Sharon, N. (1980). What should be called a lectin? *Nature* **285**(5760), 66-66.

Goodison, S., Rosser, C. J. & Urguidi, V. (2009). Urinary proteomic profiling for diagnostic bladder cancer biomarkers. *Expert Rev Proteomics* **6**(5), 507-514.

Gordon, J. N., Green, S. R. & Goggin, P.M. (2005). Cancer cachexia. *Qjm* **98**(11), 779-788.

Graves, P. R. & Haystead T. A. (2002). Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* **66**(1), 39-63.

Gupta, D. & Lis, C. G. (2009). Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *J Ovarian Res* **2**, 13.

Hakomori, S. (1985). Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. *Cancer Res* **45**(6), 2405-2414.

Hamid, R., Masood, A., Wani, I.H. & Rafiq, S. (2013). Lectin: proteins with diverse applications. *J. Applied Pharmaceutical Science* **3**(4), 93-103.

Hale, L. P., Price, D. T., Sanchez, L. M., Denmark-Wahnefried, W. & Madden, J. F. (2001). Zinc alpha-2-glycoprotein is expressed by malignant prostatic epithelium and may serve as a potential serum marker for prostate cancer. *Clin Cancer Res* **7**(4), 846-853.

Hashim, O. H., Ahmad, F. & Shuib, A. S. (2001). The application of *Artocarpus integer* seed lectin-M in the detection and isolation of selective human serum acute-phase proteins and immunoglobulins. *Immunol Invest* **30**(2), 131-141.

Hashim, O. H., Gendeh, G. S., Cheong, C. N. & Jaafar, M. I. (1994). Effect of *Artocarpus integer* lectin on functional activity of guinea-pig complement. *Immunol Invest* **23**(2), 153-160.

Hashim, O. H., Ng, C. L., Gendeh, S. & Nik Jaafar, M. I. (1991). IgA binding lectins isolated from distinct Artocarpus species demonstrate differential specificity. *Mol Immunol* **28**(4-5), 393-398.

Hassan, M. I., Waheed, A., Yaday, S., Singh, T. P. & Ahmad, F. (2008). Zinc alpha 2-glycoprotein: a multidisciplinary protein. *Mol Cancer Res* **6**(6), 892-906.

Haugen, T. H., Hanley, J. M. & Health, E. C. (1981). Haptoglobin. A novel mode of biosynthesis of a liver secretory glycoprotein. *J Biol Chem* **256**(3), 1055-1057.

Herve, F., Duche, J. C. & Jaurand, M. C. (1998). Changes in expression and microheterogeneity of the genetic variants of human alpha1-acid glycoprotein in malignant mesothelioma. *J Chromatogr B Biomed Sci Appl* **715**(1), 111-123.

Hutchens, T. W. & Yip T. T. (1993). New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Sp* **7**(7), 576-580.

Iorio, L. & Avagliano, F. (1999). Observations on the Liber medicine orinalibus by Hermogenes. *Am J. Neprol* **191**, 185-188.

Jacobs, I., Gentry-Maharaj, A., Burnell, M., Manchanda, R., Singh, N., *et al.* (2011). Sensitivity of transvaginal ultrasound screening for endometrial cancer in

postmenopausal women: a case-control study within the UKCTOCS cohort. *Lancet Oncol* **12**(1), 38-48.

Jayapalan, J. J., Ng, K. L., Razack, A. H. & Hashim, O. H. (2012). Identification of potential complementary serum biomarkers to differentiate prostate cancer from benign prostatic hyperplasia using gel- and lectin-based proteomics analyses. *Electrophoresis* **33**(12), 1855-1862.

Kaaks, R., Lukanova, A. & Kurzer, M. S. (2002). Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiol Biomarkers Prev* **11**(12), 1531-1543.

Kacinski, B. M., Carter, D., Mittal, K., Yee, L. D., Scata, K. A., *et al.* (1990). Ovarian adenocarcinomas express fms-complementary transcripts and fms antigen, often with coexpression of CSF-1. *Am J Pathol* **137**(1), 135-147.

Kang, S. M., Sung, H. J., Ahn, J. M., Park, J. Y., Lee, S. Y., *et al.* (2011). The Haptoglobin beta chain as a supportive biomarker for human lung cancers. *Mol Biosyst* **7**(4), 1167-1175.

Kapron, J. T., Hilliard, G. M., Lakins, J. N., Tenniswood, M. P., West, K., A. *et al.* (1997). Identification and characterization of glycosylation sites in human serum clusterin. *Protein Sci* **6**(10), 2120-213.

Karas, M., Bachmann, D. & Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal Chem* **57**(14), 2935-2939.

Kaur, H., Silverman, P. M., Iyer, R. B., Verschraegen, C. F., Eifel, P. J., *et al.* (2003). Diagnosis, Staging, and Surveillance of Cervical Carcinoma. *AJR Am J Roentgenol* **180**(6), 1621-1631.

Keegan, H., Mc Inerney, J., Pilkington, L., Gronn, P., Silva, I., *et al.* (2009). Comparison of HPV detection technologies: Hybrid capture 2, PreTect HPV-Proofer and analysis of HPV DNA viral load in HPV16, HPV18 and HPV33 E6/E7 mRNA positive specimens. *J Virol Methods* **155**(1), 61-66.

Kocourek, J. & Horejsi, V. (1983). A note of the recent discussion on definition of the term "lectin". *Lectins:Biology, Biochemistry, Clin Biochem* **3**, 3-6.

Kreunin, P., Zhao, J., Rosser, C., Urguidi, V., Lubman, D. M., *et al.* (2007). Bladder Cancer Associated Glycoprotein Signatures Revealed by Urinary Proteomic Profiling. *J Proteome Res* **6**(7), 2631-2639.

Laiko, V. V., Moyer, S. C. & Cotter, R. J. *et al.* (2000). Atmospheric pressure MALDI/ion trap mass spectrometry. *Anal Chem* **72**(21), 5239-5243.

Lee, Y. S., Kim, J. H., Choi, K. J., Choi, I. K., Kim, H., *et al.* (2006). Enhanced antitumor effect of oncolytic adenovirus expressing interleukin-12 and B7-1 in an immunocompetent murine model. *Clin Cancer Res* **12**(19), 5859-5868.

Lim, G. C. C. & Yahya, H. (2003). Second report of the national cancer registry cancer incidence in Malaysia. *National Cancer Registry, Ministry of Health, Kuala Lumpur*.

Lim, S. B., Chua, C. T. & Hashim, O. H. (1997). Isolation of a mannose-binding and IgE- and IgM-reactive lectin from the seeds of *Artocarpus integer*. *J Immunol Methods* **209**(2), 177-186.

Lim, S. B., Kanthimathi, M. S. & Hashim, O. H. (1998). Effect of the mannose-binding *Artocarpus integer* lectin on the cellular proliferation of murine lymphocytes. *Immunol Invest* **27**(6), 395-404.

Liu, H., Sadygov, R. G. & Yates, J. R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**(14), 4193-4201.

Liu, Y., Cao, D. J., Sainz, I. M., Guo, Y. L. & Colman, R. W. (2008). The inhibitory effect of HKa in endothelial cell tube formation is mediated by disrupting the uPA-uPAR complex and inhibiting its signaling and internalization. *Am J Physiol Cell Physiol* **295**(1), 21.

Lord, J. M., Hartley, M. R. & Roberts, L. M. (1991). Ribosome inactivating proteins of plants. *Semin Cell Biol* **2**(1), 15-22.

Lu, K. H. (2008). Hereditary gynecologic cancers: differential diagnosis, surveillance, management and surgical prophylaxis. *Fam Cancer* **7**(1), 53-58.

Madera, M., Mechref, Y., Klouckova, I. & Novotny, M. V. (2006). Semiautomated high-sensitivity profiling of human blood serum glycoproteins through lectin preconcentration and multidimensional chromatography/tandem mass spectrometry. *J Proteome Res* **5,** 2348–2363.

Marimuthu, A., O'Meally, R. N., Chaerkady, R., Subbannayya, Y., Nanjappa, V., *et al.* (2011). A comprehensive map of the human urinary proteome. *J Proteome Res* **10**(6), 2734-2743.

Menon, U. (2004). Ovarian cancer screening. *CMAJ* **171**(4), 323-324.

Mody, R., Joshi, S. & Chaney, W. (1995). Use of lectins as diagnostic and therapeutic tools for cancer. *J Pharmacol Toxicol Methods* **33**(1), 1-10.

Mohamed, E., Abdul-Rahman, P. S., Doustjalali, S. R., Chen, Y., Lim, B. K., *et al.* (2008). Lectin-based electrophoretic analysis of the expression of the 35 kDa inter-alpha-trypsin inhibitor heavy chain H4 fragment in sera of patients with five different malignancies. *Electrophoresis* **29**(12), 2645-2650.

Molden, T., Kraus, I., Skomedal, H., Nordstrom, T. & Karlsen, F. (2007). PreTect HPV-Proofer: real-time detection and typing of E6/E7 mRNA from carcinogenic human papillomaviruses. *J Virol Methods* **142**(1-2), 204-212.

Morrissey, J. J., London, A. N., Luo, J. & Kharasch, E. D. (2010). Urinary biomarkers for the early diagnosis of kidney cancer. *Mayo Clin Proc* **85**(5), 413-421.

Mu, A. K. W., Lim, B. K., Hashim, O. H. & Shuib, A. S. (2012). Detection of differential levels of proteins in the urine of patients with endometrial cancer: analysis using two-dimensional gel electrophoresis and *O*-glycan binding lectin. *Int J Mol Sci* **13**(8), 9489-9501.

Mu, A. K. W., Lim, B. K., Hashim, O. H. & Shuib, A. S. (2013). Identification of *O*-glycosylated proteins that are aberrantly excreted in the urine of patients with early stage ovarian cancer . *Int J Mol Sci* **14**(4), 7923-7931.

Murray, K. P., Mathure, S., Kaul, R., Khan, S. & Carson, L. F. (2000). Expression of complement regulatory proteins-CD 35, CD 46, CD 55, and CD 59-in benign and malignant endometrial tissue. *Gynecol Oncol* **76**(2), 176-182.

Nafee, A. M., Pasha, H. F., Abd El Aal, S. M. & Mostafa, N. A. (2012). Clinical significance of serum clusterin as a biomarker for evaluating diagnosis and metastasis potential of viral-related hepatocellular carcinoma. *Clin Biochem* **2**, 2.

Nakajima, T., Kogure, T. & Furukawa, K. (1986). Specificity of hemagglutinin of *Falcata japonica* which reacts with blood group active N-acetyl-D-galactosamine residues. *Exp Clin Immunogenet* **3**(4), 187-194.

Odicino, F., Pecorelli, S., Zigliani, L. & Creasman, W. T. (2008). History of the FIGO cancer staging system. *Int J Gynecol Obstet* **101**, 205-210.

Oh, J., Pyo, J. H., Jo, E. H., Hwang, S. I., Kang, S. C., *et al.* (2004). Establishment of a near-standard two-dimensional human urine proteomic map. *Proteomics* **4**(11), 3485-3497.

Omar, Z. A., Ali, Z. M. & Tamin, N. S. I. (2006). Malaysian cancer statistics-data and figure peninsular Malaysia 2006 national. *National Cancer Registry, Ministry of Health Malaysia.*

Ottenheijm, C. A. C. & Granzier, H. (2010). Lifting the nebula: Novel insights into skeletal muscle contractility. *Physiology* **25**, 304–310.

Pappas, C. T., Bliss, K. T., Zieseniss, A. & Gregorio, C. C. (2011). The nebulin family: An actin support group. *Trends Cell Biol* **21**, 29–37.

Paprocka, M., Wiedlocha, A., Walzel, H. & Radzikowski, C. (1992). The activity of two immunotoxins composed of monoclonal antibody MoAb-16 and A-chain of ricin (MoAb-16-RTA) or A-chain of mistletoe lectin I (MoAb-16-MLIA). *Arch Immunol Ther Exp* **40**(3-4), 223-227.

Park, J. Y., Kim, S. A., Chung, J. W., Bang, S., Park, S. W. *et al.* (2011). Proteomic analysis of pancreatic juice for the identification of biomarkers of pancreatic cancer. *J Cancer Res Clin Oncol* **137**, 1229–1238.

Pedersen, L. M. & Milman, N. (1996). Prevalence and prognostic significance of proteinuria in patients with lung cancer. *Acta Oncol* **35**(6), 691-695.

Peracaula, R., Barrabes, S., Sarrats, A., Rudd, P. M. & de Llorens, R. (2008). Altered glycosylation in tumours focused to cancer diagnosis. *Dis Markers* **25**(4-5), 207-218.

Petri, A. L., Simonsen, A. H., Yip, T. T., Hogdall, E., Fung, E. T., *et al.* (2009). Three new potential ovarian cancer biomarkers detected in human urine with equalizer bead technology. *Acta Obstet Gynecol Scand* **88**(1), 18-26.

Peumans, W. & Van Damme E. (1995). The role of lectins in plant defence. *The Histochemical Journal* **27**(4), 253-271.

Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M. *et al.* (2004). Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* **4**(4), 1159-1174.

Poljak, M., Marin, I. J. Seme, K. & Vince, A. (2002). Hybrid Capture II HPV Test detects at least 15 human papillomavirus genotypes not included in its current high-risk probe cocktail. *J Clin Virol* **25**(3), S89-97.

Pradhan, M., Risberg, B. A., Trope, C. G., van de Rijn, M., Gilks, C. B., *et al.* (2010). Gross genomic alterations and gene expression profiles of high- grade serous carcinoma of the ovary with and without BRCA1 inactivation. *BMC Cancer* **10,** 493.

Purdie, D. M. (2003). Epidemiology of endometrial cancer. *Reviews in Gynaecological Practice* **3**(4), 217-220.

Ratnam, S., Coutlee, F., Fontaine, D., Bentley, J., Escott, N. *et al.* (2010). Clinical performance of the PreTect HPV-Proofer E6/E7 mRNA assay in comparison with that of the Hybrid Capture 2 test for identification of women at risk of cervical cancer. *J Clin Microbiol* **48**(8), 2779-2785.

Reddington, J.J., Reddington, G.M. & MacLachlan, N.J. (1991). A competitive ELISA for detection of antibodies to the group antigen of bluetongue virus. *J Vet Diagn Invest* **3**, 144-147.

Redondo, M., Villar, E., Torres-Munoz, J., Tellez, T., Morell, M., *et al.* (2000). Overexpression of clusterin in human breast carcinoma. *Am J Pathol* **157**(2), 393-399.

Reed, J. C., Alpers, J. D., Nowell, P. C. & Hoover, R. G. (1986). Sequential expression of protooncogenes during lectin-stimulated mitogenesis of normal human lymphocytes. *Proc Natl Acad Sci U S A* **83**(11), 3982-3986.

Reed, N., Millan, D., Verheijen, R., Castiglione, M. & ESMO Guidelines Working Group (2010). Non-epithelial ovarian cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **21**(5), v31-v36.

Renkonen, K. O. (1948). Studies on hemagglutinin present in some representatives of the family of Leguminosae. *Ann Med Exp Biol Fenn* **26**, 66-72.

Rizzi, F. & Bettuzzi S. (2009). Clusterin (CLU) and prostate cancer. *Adv Cancer Res* **105**, 1-19.

Rodriguez-Pineiro, A. M., Ayude, D., Rodriguez-Berrocal, F. J. & Paez de la Cadena, M. (2004). Concanavalin A chromatography coupled to two-dimensional gel electrophoresis improves protein expression studies of the serum proteome. *J Chromatogr B Analyt Technol Biomed Life Sci* **803**(2), 337-343.

Roeise, O., Sivertsen, S., Ruud, T. E., Bouma, B. N., Stadaas, J. O., *et al*. (1990). Studies on components of the contact phase system in patients with advanced gastrointestinal cancer. *Cancer* **65**(6), 1355-1359.

Rogers, M. A., Clarke, P., Noble, J., Munro, N. P., Paul, A., *et al.* (2003). Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer Res* **63**(20), 6971-6983.

Rosenberg, M. E. & Silkensen, J. (1995). Clusterin: physiologic and pathophysiologic considerations. *Int J Biochem Cell Biol* **27**(7), 633-645.

Rudman, D., Chawla, R. K. & Nixon, D. W. (1978). The specific proteinuria of cancer patients. *Trans Assoc Am Physicians* **91**, 229-241.

Saldova, R., Royle, L., Radcliffe, C. M., Abd Hamid, U. M., Evans, R., *et al.* (2007). Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiol* **17**(12), 1344-1356.

Schall, W. O., Whittington, W. L., Rice, R. J. & Larsen, S. A. (1985). Epidemiological characterization of *Neisseria gonorrhoeae* by lectins. *J. Clin. Microbiol* **22**(3):379-382.

Schaub, S., Rush, D., Wilkins, J., Gibson, I. W., Weiler, T., *et al.* (2004a). Proteomic-based detection of urine proteins associated with acute renal allograft rejection. *J Am Soc Nephrol* **15**(1), 219-227.

Schaub, S., Wilkins, J., Weiler, T., Sangster, K., Rush, D., *et al.* (2004b). Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney Int* **65**(1), 323-332.

Schoen, R. E., Weissfeld, J. L. & Kuller, L. H. (1994). Are women with breast, endometrial, or ovarian cancer at increased risk for colorectal cancer? *Am J Gastroenterol* **89**(6), 835-842.

Senba, M. & Mori, N. (2012). Mechanisms of virus immune evasion lead to development from chronic inflammation to cancer formation associated with human papillomavirus infection.

Seriramalu, R., Pang, W. W., Jayapalan, J. J., Mohamed, E., Abdul-Rahman, P. S., *et al*. (2010). Application of champedak mannose-binding lectin in the glycoproteomic profiling of serum samples unmasks reduced expression of alpha-2 macroglobulin and complement factor B in patients with nasopharyngeal carcinoma. *Electrophoresis* **31**(14), 2388-2395.

Shannan, B., Seifert, M., Boothman, D. A., Tilgen, W. & Reichrath, J. (2006). Clusterin and DNA repair: a new function in cancer for a key player in apoptosis and cell cycle control. *J Mol Histol* **37**(5-7), 183-188.

Sharon, N. & Lis, H. (2004). History of lectins: from hemagglutinins to biological recognition molecules. *Glycobio* **14**(11), 30.

Sherman, M. E., Sturgeon, S., Brinton, L. & Kurman, R. J. (1995). Endometrial cancer chemoprevention: implications of diverse pathways of carcinogenesis. *J Cell Biochem Suppl* **23**, 160-164.

Shih Ie, M. & Kurman, R. J. (2004). Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am J Pathol* **164**(5), 1511-1518.

Soslow, R. A. (2008). Histologic subtypes of ovarian carcinoma: an overview. I*nt J Gynecol Pathol* **27**(2), 161-174.

Sumi, S., Arai, K. & Yoshida, K. (2001). Separation methods applicable to prostate cancer diagnosis and monitoring therapy. *J Chromatogr B Biomed Sci Appl* **764**(1-2), 445-455.

Sumner, J.B. & Howell, S.F. (1936). Identification of hemagglutinin of Jack bean with Concanavalin A. *J. Bacteriol* **32**(2), 227-237.

Sun, W., Li, F., Wu, S., Wang, X., Zheng, D., *et al.* (2005). Human urine proteome analysis by three separation approaches. *Proteomics* **5**(18), 4994-5001.

Sungur, C. (1995). Renal Cell Carcinoma and Erythropoietin. *Ann. Intern. Med* **123**(9), 732-733.

Takagaki, Y., Kitamura, N. & Nakanishi, S. (1985). Cloning and sequence analysis of cDNAs for human high molecular weight and low molecular weight prekininogens. Primary structures of two human prekininogens. *J Biol Chem* **260**(14), 8601-8609.

Tan, L. B., Chen, K. T., Yuan, Y. C., Liao, P. C. & Guo, H. R. (2010). Identification of urine PLK2 as a marker of bladder tumors by proteomic analysis. *World J Urol* **28**(1), 117-122.

Tanaka, Y., Akiyama, H., Kuroda, T., Jung, G., Tanahashi, K. *et al.*(2006). A novel approach and protocol for discovering extremely low-abundance proteins in serum. *Proteomics* **6**, 4845–4855.

Tantipaiboonwong, P., Sinchaikul, S., Sriyam, S., Phutrakul, S. & Chen, S. T. (2005). Different techniques for urinary protein analysis of normal and lung cancer patients. *Proteomics* **5**(4), 1140-1149.

Theodorescu, D., Schiffer, E., Bauer, H. W., Douwes, F., Eichhorn, F., *et al.* (2008). Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics Clin Appl* **2**(4), 556-570.

Thongboonkerd, V., Klein, J. B., Jevans, A. W. & McLeish, K. R. (2004). Urinary proteomics and biomarker discovery for glomerular diseases. *Contrib Nephrol* **141**, 292-307.

Thongboonkerd, V., McLeish, K. R., Arthur, J. M. & Klein, J. B. (2002). Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney Int* **62**(4), 1461-1469.

Tonevitsky, A. G., Toptygin, A., Pfuller, U., Bushueva, T. L., Ershova, G. V., *et al.* (1991). Immunotoxin with mistletoe lectin I A-chain and ricin A-chain directed against CD5 antigen of human T-lymphocytes; comparison of efficiency and specificity. *Int J Immunopharmacol* **13**(7), 1037-1041.

Tosner, J., Krejsek, J. & Louda, B. (1988). Serum prealbumin, transferrin and alpha-1-acid glycoprotein in patients with gynecological carcinomas. *Neoplasma* **35**(4), 403-411.

Trinchieri, G., Pflanz, S. & Kastelein, R. A. (2003). The IL-12 family of heterodimeric cytokines: new players in the regulation of T cell responses. *Immunity* **19**(5), 641-644.

Trougakos, I. P. & Gonos, E. S. (2002). Clusterin/apolipoprotein J in human aging and cancer. *Int J Biochem Cell Biol* **34**(11), 1430-1448.

Ueda, K., Fukase, Y., Katagiri, T., Ishikawa, N., Irie, S., *et al.* (2009). Targeted serum glycoproteomics for the discovery of lung cancer-associated glycosylation disorders using lectin-coupled ProteinChip arrays. *Proteomics* **9**(8), 2182-2192.

Ueda, Y., Enomoto, T., Kimura, T., Miyatake, T., Yoshino, K., *et al.* (2010). Serum Biomarkers for Early Detection of Gynecologic Cancers. *Cancers* **2**(2), 1312-1327.

Van Damme, E. J., A. Barre, et al. (2004). Cytoplasmic/nuclear plant lectins: a new story. *Trends Plant Sci* **9**(10), 484-489.

Van Damme, E. J., Goossens, K., Rouge, P. & Peumans, W. J. (1995). The major tuber storage protein of araceae species is a lectin. Characterization and molecular cloning of the lectin from *Arum maculatum* L. *Plant Physiol* **107**(4), 1147-1158.

Vandenborre, G., Smagghe, G. & Van Damme, E. J. (2011). Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* **72**(13), 1538-1550.

Vivekanandan-Giri, A., Slocum, J. L., Buller, C. L., Basrur, C. L., Ju, W., *et al.* (2011). Urine glycoprotein profile reveals novel markers for chronic kidney disease. *Int J Proteomics* **214715**, 10.

Wan, C., La, Y., Zhu, H., Yang, Y., Jiang, L., *et al.* (2007). Abnormal changes of plasma acute phase proteins in schizophrenia and the relation between schizophrenia and haptoglobin (Hp) gene. *Amino Acids* **32**(1), 101-108.

Wang, L., Li, F., Sun, W., Wu, S., Wang, X., *et al.* (2006). Concanavalin A-captured glycoproteins in healthy human urine. *Mol Cell Proteomics* **5**(3), 560-562.

Wang, Y. S., Cao, R., Jin, H., Huang, Y. P., Zhang, X. Y., *et al.* (2011). Altered protein expression in serum from endometrial hyperplasia and carcinoma patients. *J Hematol Oncol* **4**, 15.

Watson, P. & Lynch, H. T. (1993). Extracolonic cancer in hereditary nonpolyposis colorectal cancer. *Cancer* **71**(3), 677-685.

Weeks, M. E., Hariharan, D., Petronijevic, L., Radon, T. P., Whiteman, H. J., *et al.* (2008). Analysis of the urine proteome in patients with pancreatic ductal adenocarcinoma. *Proteomics* **2**(7-8), 1047-1057.

Witvrouw, M., Fikkert, V., Hantson, A., Pannecougue, C., O'keefe, B. R., *et al.* (2005). Resistance of human immunodeficiency virus type 1 to the high-mannose binding agents cyanovirin N and concanavalin A. *J Virol* **79**(12), 7777-7784.

Wong, L. P., Wong, Y. L., Low, W. Y., Khoo, E. M. & Shuib, R. (2009). Knowledge and awareness of cervical cancer and screening among Malaysian women who have never had a Pap smear: a qualitative study. *Singapore Med J* **50**(1), 49-53.

Xia, T., Zheng, Z. G., Gao, Y., Mou, H. Z., Xu, S. H., *et al.* (2008). Application of SELDI-TOF serum proteome profiling in cervical squamous cell carcinoma. *Ai Zheng* **27**(3), 279-282.

Yang, N., Feng, S., Sheddeh, K., Xie, X., Liu, Y., *et al.* (2011). Urinary glycoprotein biomarker discovery for bladder cancer detection using LC/MS-MS and label-free quantification. *Clin Cancer Res* **17**(10), 3349-3359.

Young, R. H., Oliva, E. & Scully, R. E. (1994). Small cell carcinoma of the ovary, hypercalcemic type. A clinicopathological analysis of 150 cases. *Am J Surg Pathol* **18**(11), 1102-1116.

Zhang, H., Kong, B., Qu, X., Jia, L., Deng, B. & Yang, Q. (2006). Biomarker discovery for ovarian cancer using SELDI-TOF-MS. *Gynecol Oncol* **102**(1), 61-66.

Zhang, L. Y., Ying, W. T., Mao, Y. S., He, H. Z., Liu, Y., *et al.* (2003). Loss of clusterin both in serum and tissue correlates with the tumorigenesis of esophageal squamous cell carcinoma via proteomics approaches. *World J Gastroenterol* **9**(4), 650-654.

Zhao, C., Annamalai, L., Guo, C., Kothandaraman, N., Koh, S. C., *et al.* (2007). Circulating haptoglobin is an independent prognostic factor in the sera of patients with epithelial ovarian cancer. *Neoplasia* **9**(1), 1-7.

Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A. & Lubman, D. M. (2006). Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J Proteome Res* **5**(7), 1792-1802.

Zhu, L. R., Zhang, W. Y., Yu, L., Zheng, Y. H., Hu, J., *et al*. (2008). Proteomic patterns for endometrial cancer using SELDI-TOF-MS. *J Zhejiang Univ Sci B* **9**(4), 286-290.

Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., *et al.* (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **5**(9), 2339-2347.

http://www.crystbbk.ac.uk

http://reproductivesysteminfo.com

# Appendix